

Semiparametric Identification and Fisher Information

Juan Carlos Escanciano*

Indiana University

September 20th, 2016

Abstract

This paper offers a systematic approach to semiparametric identification based on statistical information in semiparametric and nonparametric models. It first establishes a generalized rank condition that is sufficient and necessary for semiparametric identification in models with densities that are affine in a possibly unidentified nonparametric parameter. “Irregularly identified” parameters, i.e. identified parameters with zero semiparametric Fisher information, are characterized in this setting. Sufficient conditions for irregular identification are given for nonlinear models in terms of a novel “generalized Fisher information”. The paper also establishes a generic zero information result, and uses it to prove that the non-random parameters of the widely used semiparametric mixed Logit model are irregularly identified. Other example applications include regular identification of average marginal effects in binary choice panel data models with fixed effects, and irregular identification of the cumulative distribution and quantiles of the distribution of unobserved heterogeneity in nonparametric structural models of unemployment duration and random coefficient models.

Keywords: Identification; Irregular Identification; Semiparametric Models; Fisher Information.

JEL classification: C14; C31; C33; C35

*Department of Economics, Indiana University, 349 Wylie Hall, 100 South Woodlawn Avenue, Bloomington, IN 47405–7104, USA. E-mail: jescanci@indiana.edu. Web Page: <http://mypage.iu.edu/~jescanci/>. I thank Whitney Newey, Jack Porter, Pedro Sant’Anna and seminar participants at Boston College, MIT and Vanderbilt for useful comments.

1 Introduction

Nonparametric identification has become the benchmark for reliable empirical analysis in economics. Unfortunately, many nonparametric economic models of practical interest are not nonparametrically point-identified under weak assumptions; see for example discrete choice models with nonparametric unobserved heterogeneity. Yet, interesting aspects of the model might be point-identified under the same weak assumptions, a situation that is henceforth referred to as *semiparametric identification*. Although this observation has long been recognized in economics (see the early discussion in Hurwicz 1950), there is currently no systematic method available to assess which aspects of an unidentified model are identifiable and which ones are not. Furthermore, even when a model is nonparametrically identified there could be many parameters of the model that are only “irregularly identified” (in the sense of being identified but having a zero semiparametric Fisher information, see e.g. Chamberlain 1986, Heckman 1990, and Khan and Tamer 2010).¹ The main goal of this paper is to establish general conditions for regular and irregular semiparametric identification. The results for irregular identification have important practical implications, as inferences on such parameters are expected to be unstable in empirical analysis. In particular, if a parameter is irregularly identified no regular estimator with a parametric rate of convergence exists (see Chamberlain 1986).

The paper first provides a necessary and sufficient condition for regular and irregular identification, respectively, in models with densities that are affine in a nonparametric parameter. The characterization of irregular identification is used to show that semiparametric irregular identification is a common feature of many economic models of practical interest. For example, it is shown that nonparametric and semiparametric structural models with conditional densities of observables given heterogeneity that are smooth in parameters indexing the nonparametric unobserved heterogeneity are canonical examples of models with many irregularly identified functionals. The forerunners of research on this class of models are Heckman and Singer (1984a, 1984b), see also Elbers and Ridder (1982), who show that without prior information on the nonparametric unobserved heterogeneity, their structural model for a single spell of non-employment is not identified. They find moment or tail restrictions on the distribution of unobserved heterogeneity that guarantee nonparametric identification. More recently and in a similar spirit, Alvarez, Borovicková and Shimer (2016) propose a structural model of unemployment duration with two spells and nonparametric heterogeneity, and show that their model is nonparametrically unidentified. These authors discuss a prior sign restriction on the nonparametric parameter that leads to nonparametric identification. Complementing their identification results, I apply the results of this paper to show that the high smoothness of the structural model implies that many functionals of the distribution of unobserved heterogeneity are only irregularly identified. In particular, functionals such as the cumulative distribution function (cdf), quantiles or other functionals with discontinuous influence functions are generally only irregularly identified in models such as those studied in Heckman and Singer (1984a, 1984b) and Alvarez, Borovicková and Shimer (2016).

¹Other terms used in the literature for “irregular identification” are “identification at infinity” and “thin set identification”, see Chamberlain (1986), Heckman (1990) and Khan and Tamer (2010) for terminology and important examples in econometrics.

The key statistical concept used in this paper is that of score operator; see e.g. Begun, Hall, Huang and Wellner (1983). The score operator is the pathwise derivative of the log density with respect to the nonparametric parameter. The information operator, which extends the Fisher information matrix to nonparametric models, is the “variance” of the score operator. Nonparametric identification generally requires a “full rank” condition for the information operator, which may not hold in many applications. For example, it is shown that for a general class of models with missing observations the score operator is a conditional mean operator and the full rank of the Fisher information operator is equivalent to the L_2 -completeness of the conditional distribution of unobservables given the observables, which generally fails when observables have smaller support than unobservables (e.g. in discrete choice models).² This link with completeness can be used to obtain a nonparametric order identification condition which implies explicit limits on the dimension of unobservables allowed for identification. I also use the conditional mean representation of scores to explicitly characterize which linear functionals are identified when observables have discrete support.

The main results are then applied to semiparametric models that are nonlinear in the parameter of interest but linear in nuisance parameters. Examples include commonly used linear and nonlinear panel data models, semiparametric random coefficient models, semiparametric structural duration models, and models for games with multiple equilibria, among many others. It is shown that under mild smoothness conditions a positive semiparametric Fisher information for the parameter implies its local identification, thereby extending the parametric results of Fisher (1966) and Rothenberg (1971). A generalization of this result including irregular identification and a new “generalized Fisher information” quantity is also established. In a large class of semiparametric models with nonparametric unobserved heterogeneity, I show that if the distribution of observables given the unobservables is L_2 -complete then the information for the parametric component is zero. Heuristically, in these models the information loss from not knowing the nonparametric unobserved heterogeneity is so big so as to make the information for the parametric part zero and its identification extremely fragile.

This generic zero information result unifies a number of existing results in the literature, such as, for example, the zero information calculations of Chamberlain (2010) for nonlinear panel data binary choice, Ishwaran (1999) for some semiparametric mixtures of exponential families, and Bajari, Hahn, Hong and Ridder (2011) for games with multiple equilibria, among others. I use this observation to establish the zero information for the non-random coefficients in the semiparametric mixed Logit (random coefficient Logit), complementing the nonparametric identification result of Fox, Kim, Ryan and Bajari (2012). This new zero information result for the mixed Logit implies that the root-n estimability often used in the applied literature depends crucially on functional form assumptions for the distribution of the random coefficients (e.g. Gaussian distributional assumptions).

The identification problem has a long history and has been extensively investigated in econometrics, see the earlier studies by Koopmans (1949), Hurwicz (1950), Koopmans and Reiersol (1950), Fisher (1966) and Rothenberg (1971), among others. Bekker and Wansbeek (2001) and Dufour and Liang (2014) discuss more recent contributions and provide a survey of existing results in parametric set-

²This completeness is related but different from completeness in nonparametric instrumental variables, see Newey and Powell (2003), where completeness is between observables (endogenous variables and instruments).

tings. In semiparametric and nonparametric models, Chamberlain (1986) and Van der Vaart (1991) investigate the related question of regular estimation (or lack thereof). Van der Vaart (1991) shows that certain differentiability of a functional is necessary for regular estimation, and sufficient for the semiparametric identification condition used here, but he did not prove identification results, which are the main objective of this paper. Chamberlain’s (1986) impossibility theorems allowed for lack of identification of the nonparametric model, which is also permitted here. This level of generality is needed if one wants to include important applications such as discrete choice models with nonparametric heterogeneity. See also Newey (1990) for other impossibility theorems. There are of course many papers reporting sufficient conditions for identification in specific nonparametric models, see e.g. the reviews in Matzkin (2007, 2013) and Lewbel (2016) and references therein. Recently, Chen, Chernozhukov, Lee and Newey (2014) have provided sufficient conditions for local identification of the nonparametric parameter and regular semiparametric identification for conditional moment restriction models. Chen and Santos (2015) investigate local nonparametric regular overidentification. The present paper is particularly concerned with “irregular semiparametric identification”, where the semiparametric information is zero for the parameter of interest, and the full nonparametric parameter may not be identified. Irregular identification is also the main theme of Khan and Tamer (2010), who provide important examples and investigate its impact on estimation; see also Chen and Pouzo (2015) for general inference results for irregular functionals.³ Also related is the identification analysis of Severini and Tripathi (2006, 2012) in nonparametric instrumental variables models. They characterized the set of linear continuous functionals that are regularly and irregularly identified when the nonparametric structural regression is not necessarily identified. Santos (2011) and Davezies (2015) investigate regular estimation in this setting. The present paper deals broadly with semiparametric identification in nonparametric/semiparametric likelihood models, and more importantly, follows the tradition of Rothenberg (1971) in linking the identification problem to the concept of statistical information, albeit in a nonparametric setting.

The rest of the paper is organized as follows. Section 2 introduces the setting and some examples that will be used throughout. Section 3 characterizes identification, regular and irregular, in models that are affine in a nonparametric parameter. Section 4 deals with the important class of Information Loss models (see Le Cam and Yang, 1988). These are models where the observable dependent variables are a known measurable function of other observables and unobservables of arbitrary dimension. Semiparametric models are analyzed in Section 5, with a particular emphasis on models that are nonlinear in the parameter interest but linear in the nuisance parameter. Section 6 discusses two further examples in detail: the semiparametric random coefficients Logit model and a static binary choice panel data model with fixed effects. Section 7 concludes. An Appendix contains proofs of the main results and

³Notable examples of irregularly identified parameters in econometrics include, among others, densities, regression functions, and their derivatives evaluated at fixed points; regression discontinuity parameters, see Cattaneo and Escanciano (2017) for a recent review; binary choice coefficients under Manski’s (1975) conditions, see Chamberlain (1986, 2010); the intercept in a sample selection model, see Heckman (1990) and Andrews and Schafgans (1998); parameters in the mixed proportional model, see Hahn (1994) and Ridder and Woutersen (2003); or interaction parameters in triangular and simultaneous systems with binary variables, see Khan and Nekipelov (2016).

further discussion on identification conditions for general linear and nonlinear models.

2 Setting and Examples

This section introduces notation, defines formally identification, and provides example applications. The data is a random sample Z_1, \dots, Z_n from a distribution \mathbb{P} that belongs to a class of probability measures $\mathcal{P} = \{\mathbb{P}_\lambda : \lambda \in \Lambda\}$, where Λ is a subset of a Hilbert space $(\mathbf{H}, \langle \cdot, \cdot \rangle_{\mathbf{H}})$, with inner product $\langle \cdot, \cdot \rangle_{\mathbf{H}}$ and norm $\|\cdot\|_{\mathbf{H}}$. For example, in parametric models Λ is a subset of the Euclidean Space \mathbb{R}^m and $\|\cdot\|_{\mathbf{H}} = |\cdot|$ is the Euclidean norm. This paper focusses on nonparametric models where Λ is infinite-dimensional, e.g. a subset of a space of square-root probability densities. The nonparametric parameter that generated the data is denoted by $\lambda_0 \in \Lambda$, i.e. $\mathbb{P} = \mathbb{P}_{\lambda_0}$. The goal is to find sufficient conditions for identification of $\phi(\lambda)$ at $\phi(\lambda_0)$, for a functional $\phi(\lambda) : \Lambda \mapsto \mathbb{R}^p$, allowing for the full model \mathbb{P}_λ to be unidentified at λ_0 . That is, the equation $\mathbb{P}_\lambda = \mathbb{P}$ may have more than one solution in Λ . This setting includes as a special case semiparametric models where $\lambda = (\theta, \eta) \in \Lambda = \Theta \times H$, $\Theta \subset \mathbb{R}^p$ and H is a subset of another Hilbert Space $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$. In semiparametric models a leading example of functional is the finite-dimensional parameter, i.e. $\phi(\theta, \eta) = \theta$. However, the setting also includes functionals of the nuisance parameter $\phi(\lambda) = \chi(\eta)$, where $\chi : H \mapsto \mathbb{R}^p$, which, despite the name, may be of interest. For example, η can measure unobserved heterogeneity, and one might be interested in some average marginal effects or policy counterfactuals that involve averaging across a heterogeneous population.

I introduce the definition of identification. Let f_λ be the density of \mathbb{P}_λ with respect to a σ -finite measure μ . Denote by $\mathcal{B}_\delta(\lambda_0) := \{\lambda \in \Lambda : \|\lambda - \lambda_0\|_{\mathbf{H}} < \delta\}$ an open ball of radius $\delta > 0$ around λ_0 .

Definition 2.1 (Semiparametric Identification) *$\phi(\lambda)$ is locally identified in \mathcal{P} at $\phi(\lambda_0)$ if there exists $\delta > 0$ such that for all $\lambda \in \mathcal{B}_\delta(\lambda_0)$, $f_\lambda = f_{\lambda_0}$ μ -almost surely (μ -a.s.) implies $\phi(\lambda) = \phi(\lambda_0)$. If this implication holds for all $\lambda \in \Lambda$, then $\phi(\lambda)$ is (globally) identified at $\phi(\lambda_0)$.*

To simplify the exposition, I simply say “ $\phi(\lambda_0)$ is locally identified” rather than “ $\phi(\lambda)$ is locally identified in \mathcal{P} at $\phi(\lambda_0)$ ”, and if “locally” is dropped then identification is meant to be global. Existence of densities is not required for identification, but given the interest here in linking identification and statistical information, I use them throughout. For parametric models, i.e. $\Lambda \subset \mathbb{R}^m$, Fisher (1966) and Rothenberg (1971) obtained sufficient (and necessary) conditions for local identification of λ_0 . Specifically, non-singularity of the classical Fisher information matrix and sufficient “smoothness” of the model suffice for local identification of λ_0 . This paper provides sufficient and necessary conditions for semiparametric identification, i.e. identification of finite-dimensional parameters in nonparametric models. Semiparametric identification is clearly weaker than nonparametric identification. The following examples will be used to illustrate the main results.

Example 1: Unemployment Duration with Heterogeneity. Alvarez, Borovicková and Shimer (2016) propose a structural model for transitions in and out of employment that implies a duration of unemployment given by the first passage time of a Brownian motion with drift, a random variable with an inverse Gaussian distribution. The parameters of the inverse Gaussian distribution are allowed to

vary in arbitrary ways to account for unobserved heterogeneity in workers. These authors investigate nonparametric identification of the distribution of unobserved heterogeneity, denoted by $G_0(\alpha, \beta)$, when two unemployment spells $Z_i = (t_{i1}, t_{i2})$ are observed on the set \mathcal{T}^2 , $\mathcal{T} \subseteq [0, \infty)$. The reduced form parameters $(\alpha, \beta) \in \mathbb{R} \times (0, \infty)$ are functions of structural parameters. Here λ_0 is the density associated to G_0 with respect to a σ -finite measure π . The distribution of Z_i is absolutely continuous with Lebesgue density $f_{\lambda_0}(t_1, t_2)$ given by

$$f_{\lambda_0}(t_1, t_2) = \int_{\mathbb{R}^2} f_{z/\alpha, \beta}(t_1, t_2) \lambda_0(\alpha, \beta) d\pi(\alpha, \beta),$$

where the conditional density of Z given (α, β) is

$$f_{z/\alpha, \beta}(t_1, t_2) \propto f(t_1; \alpha, \beta) f(t_2; \alpha, \beta),$$

(\propto denotes equality up to multiplication by a normalizing constant) with $f(t; \alpha, \beta)$ denoting the inverse Gaussian density

$$f(t; \alpha, \beta) \propto \frac{\beta}{t^{3/2}} e^{-\frac{(\alpha t - \beta)^2}{2t}}.$$

Alvarez, Borovicková and Shimer (2016) show that λ_0 is only nonparametrically identified up to the sign of α . I complement their analysis with an investigation of semiparametric identification, i.e. which linear functionals $\phi(\lambda_0)$ are regularly or irregularly identified. In particular, this paper shows that the cdf of λ_0 at a point, and other functionals of λ_0 with discontinuous influence functions, such as quantiles, will be only irregularly identified when the nonparametric parameter λ_0 is identified. \blacktriangle

Example 2: Random Coefficients Discrete Choice. Ichimura and Thompson (1998) and Gautier and Kitamura (2013) have investigated nonparametric identification and estimation of the binary choice random coefficient model given by

$$Y_i = 1 \left(X_i' \beta_i \geq 0 \right),$$

where we observe $Z_i = (Y_i, X_i)$ but β_i is unobservable, and where $1(A)$ denotes the indicator function of the event A . The random vector β_i is independent of X_i , normalized to $|\beta_i| = 1$ and satisfies $\Pr(\beta_i = 0) = 0$. Let λ_0 denote the density of β_i with respect to the uniform spherical measure $\sigma(\cdot)$ in \mathbb{S}^{d-1} , where $\mathbb{S}^{d-1} = \{b \in \mathbb{R}^d : |b| = 1\}$ denotes the unit sphere in \mathbb{R}^d . The density of the data for a positive outcome is given by

$$f_{\lambda_0}(x) = \int_{\mathbb{S}^{d-1}} 1(x's \geq 0) \lambda_0(s) d\sigma(s).$$

Ichimura and Thompson (1998) and Gautier and Kitamura (2013) have shown that without further conditions λ_0 is generally not identified. I provide below necessary and sufficient conditions for regular or irregular identification of a linear functional $\phi(\lambda_0)$. More generally, the results of this paper can be used to investigate semiparametric identification in linear and nonlinear semiparametric random coefficient models. I illustrate the applicability of the results of this paper with the popular random coefficients Logit model used extensively in discrete choice. Specifically, I show in Section 6 that the non-random parameters are irregularly identified in the general model analyzed in Fox et al. (2012), where the distribution of random slopes is fully nonparametric. \blacktriangle

Example 3: Nonlinear Fixed Effects Panel Data. Denote the conditional density of the vector $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{iT})$ given the vector of exogenous variables $X_i = (X_{i1}, X_{i2}, \dots, X_{iT})$ and the vector of individual effects α_i by $f_{y/x,\alpha}(\cdot; \theta_0)$, which is known up to the parameter $\theta_0 \in \Theta \subset \mathbb{R}^p$. Here T is fixed. The sequence $\{(Y_i, X_i, \alpha_i)\}_{i=1}^n$ are independent and identically distributed (*iid*) draws, but only $Z_i = (Y_i, X_i)$ is observed. Let $\eta_0(\alpha, x)$ denote the unrestricted conditional Lebesgue density of α given $X = x$, which is supported on \mathcal{A} for all x . Then, the density of the observation $Z_i = (Y_i, X_i)$ is

$$f_{\lambda_0}(y, x) = \int_{\mathcal{A}} f_{y/x,\alpha}(y; \theta_0) \eta_0(\alpha, x) d\alpha,$$

where $\lambda_0 = (\theta_0, \eta_0)$. One functional of interest is the parameter θ_0 , i.e. $\phi(\theta, \eta) = \theta$. Other parameters of interest are Average Marginal Effects (AME), of the form

$$\phi(\lambda_0) = \mathbb{E}[m(\alpha_i, X_i)].$$

When the support of Y_i is discrete, λ_0 is generally not identified. Nevertheless, $\phi(\lambda_0)$ may still be identified. I provide below sufficient and necessary conditions for identification of $\phi(\lambda_0)$. There is of course an extensive literature on identification and estimation of nonlinear panel data models (see Arellano and Bohomme 2011 for a recent survey). Important identification and information results are given in Chamberlain (2010), who showed that with bounded covariates in the static binary choice model θ_0 is only identified for the Logit. When the errors are not logistic and covariates have unbounded support, Chamberlain (2010) shows that the semiparametric information bound for θ_0 is zero, but θ_0 may still be point-identified (see Manski 1987, Honoré and Kyriazidou 2000, and Hahn 2001). Bonhomme (2012) proposes a general method to estimate θ_0 under regular identification; whereas Bonhomme (2011) investigates identification and estimation of AME when θ_0 is known. After discussing Chamberlain's (2010) identification results for the binary choice panel data model in terms of the completeness condition of this paper, I discuss in Section 6 the identification of AME when θ_0 is identified but unknown, and when θ_0 is unidentified (as when covariates have bounded support). In particular, I show that the set of regularly identified AME is rather small, and provide conditions under which lack of identification of θ_0 has no impact on the identification of AME. \blacktriangle

3 Affine Nonparametric Models

This section first introduces some notation that will be used throughout the paper. For a generic measure ν , let $L_q(\nu)$, $q \geq 1$, denote the Banach space of (equivalence classes of) real-valued measurable functions h such that $\|h\|_{q,\nu} := (\int |h|^q d\nu)^{1/q} < \infty$ (henceforth I drop the sets of integration in integrals and the qualification ν -almost surely for simplicity of notation). The notation $L_q^0(\nu)$ denotes the set of functions in $L_q(\nu)$ with zero ν -mean. Define the Hilbert space L_2 , of \mathbb{P} -square integrable measurable functions with inner product $\langle h, f \rangle = \int h f d\mathbb{P}$ and norm $\|h\|^2 = \langle h, h \rangle$ (I drop the dependence on $q = 2$ and $\nu = \mathbb{P}$ in this case). The set L_2^0 is the subspace of zero mean functions in L_2 . Henceforth, for a generic linear operator $K : \mathcal{G}_1 \rightarrow \mathcal{G}_2$, $\mathcal{N}(K) := \{f \in \mathcal{G}_1 : Kf = 0\}$ denotes its kernel. Set $B_0 = \{b \in \mathbf{H} : \lambda_0 + b \in \Lambda\}$ and let $T(\lambda_0)$ denote the linear span of elements in B_0 .

Following the tradition in Rothenberg (1971), I aim to relate identification with the concept of statistical information. To that end, I consider the score operator, which under the conditions of this section is defined as the operator $S : T(\lambda_0) \mapsto L_2$ given by

$$Sb \equiv S_{\lambda_0} b := \frac{f_{\lambda_0+b} - f_{\lambda_0}}{f_{\lambda_0}} 1(f_{\lambda_0} > 0). \quad (1)$$

Intuitively, the expression for S comes from the approximation $\partial \log f_\lambda / \partial \lambda \approx (f_\lambda - f_{\lambda_0}) / f_{\lambda_0}$ for $\lambda \approx \lambda_0$. The score associated to the parametric submodel f_{λ_0+tb} at $t = 0$ is Sb , in most cases $\partial \log f_{\lambda_0+tb} / \partial t|_{t=0} = Sb$. More generally, existence of score operators is necessary for the classical mean square differentiability assumption, which means that for every path $\lambda_t \in \Lambda$ with $t^{-1}(\lambda_t - \lambda_0) \rightarrow b \in T(\lambda_0) \subset \mathbf{H}$ the following holds, as $t \downarrow 0$,

$$\left\| \frac{f_{\lambda_t}^{1/2} - f_{\lambda_0}^{1/2}}{t} - \frac{1}{2} Sb f_{\lambda_0}^{1/2} \right\|_{2,\mu} \rightarrow 0. \quad (2)$$

This section investigates identification when the model's density f_λ and the functional ϕ are affine. To simplify the exposition I assume in what follows that the functional is a scalar, with the understanding that all the results below have straightforward extensions to multivariate functionals.

Assumption 1: (i) The map $\dot{\phi} : T(\lambda_0) \subseteq \mathbf{H} \mapsto \mathbb{R}$ defined by $\dot{\phi}(b) = \phi(\lambda_0 + b) - \phi(\lambda_0)$ is linear; (ii) the score operator $S : T(\lambda_0) \subseteq \mathbf{H} \mapsto L_2$ in (1) is well defined and linear; (iii) for each $b \in \mathcal{N}(S)$, $b \neq 0$, there exists $c \equiv c(b) \in \mathbb{R}$, such that $\lambda_0 + cb \in \Lambda$.

Assumption 1(i) holds for the leading example of the finite-dimensional parameter in a semiparametric model, i.e. $\phi(\theta, \eta) = \theta$. When ϕ is linear, $\dot{\phi} = \phi$. Section 8.4 in the Appendix relaxes the linearity of $\dot{\phi}$. Assumption 1(ii) certainly holds in Examples 1 and 2 above, and in Example 3 after fixing the finite-dimensional parameter θ . In other models this assumption holds after a suitable reparametrization. For example, consider a simple model of a binary outcome Y^* that is only observed when $D = 1$. That is, the available data is $Z = (Y, D, X)$, where $Y = Y^*D$ and Y^* is independent of D given X . Define $q(x) = \mathbb{E}[Y^* | X]$ and $p(x) = \mathbb{E}[D | X]$. The density of (Y, D, X) with respect to a suitable measure is a nonlinear function of p and q , since for example $\mathbb{P}(Y = 1, D = 1 | X = x) = q(x)p(x)$, but if we reparametrize the density in terms of $\lambda_0 = (\lambda_{01}, \lambda_{02})$ with $\lambda_{01}(x) = q(x)p(x)$ and $\lambda_{02}(x) = p(x)$, then the density becomes affine in λ_0 . Section 5 below relaxes Assumption 1(ii); see also the Appendix for general results for nonlinear models. Assumption 1(iii) holds if for example λ_0 belongs to the interior of Λ . When Λ is a subset of densities, as in Examples 1 and 2, Assumption 1(iii) holds under mild conditions as discussed below. This assumption is used to prove the necessity of the main identification condition below. Overall, Assumption 1 is convenient because under this assumption identification can be characterized. Thus, the results under Assumption 1 provide a benchmark for what can be achieved in more complicated models.

In this setting, identification of $\phi(\lambda_0)$ will hold if, for all $b \in T(\lambda_0)$,

$$f_{\lambda_0+b} - f_{\lambda_0} \equiv f_{\lambda_0} 1(f_{\lambda_0} > 0) Sb = 0 \implies \phi(\lambda_0 + b) - \phi(\lambda_0) \equiv \dot{\phi}(b) = 0,$$

or, since $f_{\lambda_0} > 0$ \mathbb{P} -a.s.,

$$\mathcal{N}(S) \subset \mathcal{N}(\dot{\phi}). \quad (3)$$

Equation (3) appeared first in Van der Vaart (1991) as a necessary condition for differentiability of a functional, although no identification results were provided there. The following theorem proves that (3) is necessary and sufficient for semiparametric identification of $\phi(\lambda_0)$ under Assumption 1.

Theorem 3.1 *Let Assumption 1(i-ii) hold. Then, (3) is a sufficient condition for identification of $\phi(\lambda_0)$. If in addition Assumption 1(iii) holds, then (3) is also necessary for identification of $\phi(\lambda_0)$.*

Remark 3.1 *Note that Theorem 3.1 does not require neither the continuity of S and $\dot{\phi}$, nor the Hilbert space structure of the parameter space. This theorem holds more generally for discontinuous functionals, such as a density λ_0 evaluated at a fixed point, and/or any subset Λ of a metric space. For example, when Λ is a subset of densities with respect to π , one can replace \mathbf{H} by $L_1(\pi)$ in Theorem 3.1. Similarly, the range of S can be defined in $L_1(\mathbb{P})$. Similar extensions are also possible in some of the subsequent identification results, but for simplicity of exposition I consider a Hilbert space setting for the most part.*

The following examples illustrate the utility of Theorem 3.1 and Remark 3.1.

Example 1: Unemployment Duration with Heterogeneity, cont. In their Theorem 1, Alvarez, Borovicková and Shimer (2016) show that if $\alpha \geq 0$ or $\alpha \leq 0$ π -a.s. then λ_0 is identified. As discussed by these authors, assuming either case imposes substantial restrictions on the economic model. For example, $\alpha \geq 0$ implies that all workers return to work eventually. Here, I build on their nonparametric (un-)identification result and characterize the set $\mathcal{N}(S)$. This is useful because we can then use Theorem 3.1 to systematically see which functionals are identified and which ones are not. By Remark 3.1, the score operator S can be defined on $T(\lambda_0) \subseteq L_1(\pi)$, with values in $L_1(\mathbb{P})$, and it is given by

$$Sb = \frac{1}{f_{\lambda_0}(t_1, t_2)} \int f_{z/\alpha, \beta}(t_1, t_2) b(\alpha, \beta) d\pi(\alpha, \beta).$$

The following proposition characterizes $\mathcal{N}(S)$.

Proposition 3.2 *Assume $d\pi(-\alpha, \beta) = -d\pi(\alpha, \beta)$ and let $\mathcal{T} \subseteq [0, \infty)$ have a non-empty interior. Then,*

$$\mathcal{N}(S) = \left\{ b \in T(\lambda_0) : b(\alpha, \beta) = e^{-4\alpha\beta} b(-\alpha, \beta) \right\}.$$

A corollary of this proposition is that the identified set for λ_0 is the set of $\lambda = \lambda_0 + b \in \Lambda$, where $b \in \mathcal{N}(S)$. That is, by definition of S , $f_{\lambda_0+b} = f_{\lambda_0}$, and thus $\lambda_0 + b$ and λ_0 are observationally equivalent, if $b \in \mathcal{N}(S)$. Note that there is partial identification, i.e. the set $\mathcal{N}(S) \neq \{0\}$. Indeed, $\mathcal{N}(S)$ contains all functions in $L_1(\pi)$ of the form $b(\alpha, \beta) = C(\alpha, \beta) / (1 - e^{4\alpha\beta})$ where $C(\alpha, \beta)$ is an odd function of α . In particular, Theorem 3.1 implies that the mean of α is not identified, since for example, the function

$$b(\alpha, \beta) = \frac{\alpha}{1 - e^{4\alpha\beta}} 1(-c \leq \alpha \leq c),$$

for a positive constant c , is such that $b \in \mathcal{N}(S)$ but $\phi(b) \neq 0$ for $\phi(\lambda_0) = \mathbb{E}_{\lambda_0}[\alpha] = \int \alpha \lambda_0(\alpha, \beta) d\pi(\alpha, \beta)$. That is, the identification condition (3) does not hold for the mean of α . One assumption that leads to nonparametric identification is $\alpha \geq 0$ or $\alpha \leq 0$ π -a.s. (Alvarez et al., 2016), since then $\mathcal{N}(S) = \{0\}$. Another assumption is that Λ only contains densities that are symmetric in α , i.e. for all $\lambda \in \Lambda$, $\lambda(\alpha, \beta) = \lambda(-\alpha, \beta)$ π -a.s. Other prior restrictions could be entertained, and they will restrict $\mathcal{N}(S)$ above and consequently, the identified set. \blacktriangle

Example 2: Random Coefficients Discrete Choice, cont. In the binary choice random coefficient model the score operator $S : T(\lambda_0) \subseteq L_2(\sigma) \mapsto L_2$ is defined as

$$Sb(x) = \frac{1(f_{\lambda_0}(x) > 0)}{f_{\lambda_0}(x)} \int_{\mathbb{S}^{d-1}} 1(x's \geq 0) b(s) d\sigma(s).$$

Then,

$$\mathcal{N}(S) = \left\{ b \in L_2(\sigma) : \int_{\mathbb{S}^{d-1}} 1(x's \geq 0) b(s) d\sigma(s) = 0 \right\}. \quad (4)$$

When the support of X is \mathbb{R}^d and the distribution of X is absolutely continuous with respect to σ , Lemma 2.3 in Rubin (1999), see also Corollary A.1 in Gautier and Kitamura (2013), characterizes $\mathcal{N}(S)$ as

$$\mathcal{N}(S) = \{b \in L_2^0(\sigma) : b(s) = b(-s) \sigma - a.s.\}.$$

Thus, even with full support of X , λ_0 is not identified. Nevertheless, the mean of the random coefficients is a functional that is identified by Theorem 3.1, since condition (3) holds for that functional. Hoderlein and Sherman (2015) provide an alternative set of assumptions for the identification of the mean of random coefficients. Ichimura and Thompson (1993, Theorem 3) restrict further the domain of S so nonparametric identification is obtained. These restrictions involve continuous regressors with full support and a sign restriction on one of the coefficients. Below, results on regular and irregular identification are provided. \blacktriangle

I add to Assumption 1(i-iii) the following mild condition to establish the link with information.

Assumption 1: (iv) The maps $\dot{\phi} : T(\lambda_0) \subseteq \mathbf{H} \mapsto \mathbb{R}$ and $S : T(\lambda_0) \subseteq \mathbf{H} \mapsto L_2$ are continuous.

Associated to S is its adjoint operator $S^* : L_2 \mapsto T(\lambda_0)$, satisfying $\langle g, Sb \rangle = \langle S^*g, b \rangle_{\mathbf{H}}$ for all $g \in L_2$ and $b \in T(\lambda_0)$. The following definition extends the concept of statistical Fisher information matrix from the parametric to a nonparametric context.

Definition 3.1 *The (Fisher) information operator is defined as $I_{\lambda_0} := S^*S : T(\lambda_0) \mapsto T(\lambda_0)$.*

Roughly, $I_{\lambda_0}b$ measures the Fisher information of λ_0 in the direction $b \in T(\lambda_0)$, i.e. the classical Fisher information corresponding to f_{λ_0+tb} at the “true” parameter value $t = 0$. The following example illustrates that I_{λ_0} is a nonparametric extension of the classical Fisher information matrix.

Example 4: Parametric Models. Suppose $\mathcal{P} = \{\mathbb{P}_\lambda : \lambda \in \Lambda\}$ with $\Lambda \subset \mathbb{R}^m$ an open set. Then, under standard regularity conditions $Sb = b' \dot{l}_{\lambda_0}$, where \dot{l}_{λ_0} is the classic score for λ at λ_0 (in many cases

the derivative of the log-likelihood with respect to λ), and $\langle g, Sb \rangle = b' \langle g, \dot{l}_{\lambda_0} \rangle$, so that $S^*g = \langle g, \dot{l}_{\lambda_0} \rangle$. Here, $T(\lambda_0) = \mathbb{R}^m$. The information operator corresponds to $I_{\lambda_0}b := S^*Sb = I_{\lambda_0}b$, where with some abuse of notation $I_{\lambda_0} = \mathbb{E} [\dot{l}_{\lambda_0} \dot{l}_{\lambda_0}']$ denotes the standard Fisher information matrix. \blacktriangle

The identification condition (3) can be also written in terms of the information operator as

$$\mathcal{N}(I_{\lambda_0}) \subset \mathcal{N}(\dot{\phi}). \quad (5)$$

This follows because $\mathcal{N}(I_{\lambda_0}) = \mathcal{N}(S)$. It is more convenient, however, to work with score operators than information operators because the former have simpler representations (see e.g. Section 4 below). Equation (5) implies that nonparametric identification of λ_0 means non-singularity of the Fisher information I_{λ_0} in all directions, i.e. all linear functionals are identified iff $\mathcal{N}(I_{\lambda_0}) = \{0\}$ (cf. Proposition 3.5).

I now relate the identification condition (3) with the concept of semiparametric information. By the continuity in Assumption 1(iv), S and $\dot{\phi}$ are uniquely extended to $\overline{T(\lambda_0)}$, where henceforth for a generic set V , \overline{V} denotes the closure of V in the norm topology. Also by continuity, there exists $r_\phi \in \overline{T(\lambda_0)}$, called the Riesz's representer of ϕ , such that for all $b \in \overline{T(\lambda_0)}$,

$$\dot{\phi}(b) = \langle b, r_\phi \rangle_{\mathbf{H}}.$$

Let $\mathcal{R}(S^*) := \{f \in \overline{T(\lambda_0)} : \exists g \in L_2, S^*g = f\}$. Then, I show the following characterization of (3).

Proposition 3.3 *Under Assumption 1, (3) is equivalent to the generalized rank condition*

$$r_\phi \in \overline{\mathcal{R}(S^*)}. \quad (6)$$

Remark 3.2 *Condition (6) (hence (3)) is clearly weaker than the rank condition*

$$r_\phi \in \mathcal{R}(S^*). \quad (7)$$

Theorem 4.1 in van der Vaart (1991) shows that (7) is equivalent to the semiparametric Fisher information for $\phi(\lambda)$ at $\phi(\lambda_0)$ being positive, i.e. $I_\phi > 0$, where

$$I_\phi = \inf_{b \in \overline{T(\lambda_0)}, b \neq 0} \frac{\|Sb\|^2}{[\dot{\phi}(b)]^2}.$$

This representation of I_ϕ formalizes the intuition first provided by Stein (1956) that the semiparametric information is the infimum of the informations over all parametric submodels. Indeed, the information for estimating the parameter $\psi(t) = \phi(\lambda_t)$ under the density f_{λ_0+tb} at $t = 0$ is, by a Delta method argument, equal to $[\partial\psi(t)/\partial t]^{-1} \|Sb\|^2 [\partial\psi(t)/\partial t]^{-1} = \|Sb\|^2 / [\dot{\phi}(b)]^2$, where all derivatives are evaluated at the “truth” $t = 0$. The semiparametric Fisher information is the infimum of the information over all such parametric submodels.

That $r_\phi \in \mathcal{R}(S^)$ implies $I_\phi > 0$ follows from the adjoint property and Cauchy-Schwarz inequality, since*

$$[\dot{\phi}(b)]^2 = \langle b, r_\phi \rangle_{\mathbf{H}}^2 = \langle b, S^*r_\phi^* \rangle_{\mathbf{H}}^2 = \langle Sb, r_\phi^* \rangle^2 \leq \|Sb\|^2 \|r_\phi^*\|^2,$$

or

$$\frac{\|Sb\|^2}{[\dot{\phi}(b)]^2} \geq \frac{1}{\|r_\phi^*\|^2} > 0,$$

for some $r_\phi^* \in L_2$. This shows that (7) implies $I_\phi > 0$. The reciprocal is also true (cf. Theorem 4.1 in Van der Vaart 1991).

The following result summarizes the discussion above. An identified parameter is regularly (irregularly) identified when its semiparametric Fisher information defined above is positive (respectively, zero).

Theorem 3.4 *Let Assumption 1 hold. Then (i) $\phi(\lambda_0)$ is regularly identified iff $r_\phi \in \mathcal{R}(S^*)$; (ii) $\phi(\lambda_0)$ is irregularly identified iff $r_\phi \in \overline{\mathcal{R}(S^*)} \setminus \mathcal{R}(S^*)$; and (iii) $\phi(\lambda_0)$ is unidentified iff $r_\phi \notin \overline{\mathcal{R}(S^*)}$.*

Part (i) of Theorem 3.4 extends the necessary and sufficient condition of Rothenberg (1971) to the semiparametric case of this section. Its proof follows from a combination of Theorem 3.1, Proposition 3.3 above and Theorem 4.1 in Van der Vaart (1991). Under Assumption 1, regular identification holds iff $I_\phi > 0$. Part (ii) relaxes the positive information and characterizes irregular identification. In (ii) the semiparametric Fisher information for $\phi(\lambda_0)$ is zero, $I_\phi = 0$, although $\phi(\lambda_0)$ is identified. In this case, Theorem 2(ii) in Chamberlain (1986) implies that there exists no regular estimator for $\phi(\lambda_0)$ converging at the parametric rate.

In the parametric case of Rothenberg (1971) $\mathcal{R}(S^*)$ is finite-dimensional, and then closed (see e.g. Kress 1999, p.5). This leads to the following

Corollary 3.1 *If $\mathcal{R}(S^*)$ is closed, irregular identification is not possible.*

Irregular identification is specific to nonparametric models. There are relatively few examples, however, of nonparametric models where $\mathcal{R}(S^*)$ is closed and hence all identified functionals are regularly identified (see e.g. the measurement error model with self-reported data in An and Hu 2012, asset pricing models studied in Chen and Ludvigson 2009, Escanciano et al. 2015 and Chen et al. 2014, and other examples considered in Section 7 of Carrasco et al. 2006). Another implication of $\mathcal{R}(S^*)$ closed is that the nonparametric information operator is “regularly positive”, meaning that for all $b \in \overline{T(\lambda_0)}$ and some positive constant C ,

$$C \|b\|_{\mathbf{H}}^2 \leq \|Sb\|^2 = \langle I_{\lambda_0} b, b \rangle_{\mathbf{H}},$$

as shown in the following Proposition.

Proposition 3.5 *Let Assumption 1 hold. Then, (i) nonparametric identification of λ_0 holds iff $\mathcal{N}(I_{\lambda_0}) = \{0\}$, and (ii) if $\mathcal{R}(S^*)$ is closed, then nonparametric identification holds iff the nonparametric information operator is regularly positive.*

Proposition 3.5 is given here for completeness. It establishes sufficient and necessary conditions for nonparametric identification under Assumption 1. General sufficient conditions for nonparametric local identification in conditional moment restrictions models are given in Chen et al. (2014). It is worth

comparing Proposition 3.5 with their results⁴. The general approach of Chen et al. (2014) restricts the parameter space to obtain nonparametric local identification in general linear and nonlinear conditional moment models. The sufficiency of Proposition 3.5(i) is the likelihood analog of their Theorem 2. When conditional moments are Frechet differentiable, they consider the parameter space (in their notation) to have tangents in $\{b : \|m'b\|^2 > C \|b\|_{\mathbf{H}}^2\}$, for the derivative m' of a conditional mean operator m and a positive constant C . In our setting, an analog that allows comparison with statistical information would be $m(\lambda) = (f_\lambda - f_{\lambda_0})/f_{\lambda_0}$, with derivative $m'(\lambda) = S(\lambda - \lambda_0)$. On the parameter space with tangents $\{b : \|Sb\|^2 > C \|b\|_{\mathbf{H}}^2\}$ the information for the nonparametric parameter is regularly positive, as in Proposition 3.5(ii). Chen et al. (2014) also consider conditions corresponding to higher order differentiability and these conditions do allow for irregular nonparametric identification.

Remark 3.3 *The identification result of Theorem 3.4 is constructive in both the regular and irregular cases. Let λ^* denote the unique minimum norm element of the set $\Lambda_0 = \{\lambda : \lambda = \lambda_0 + \mathcal{N}(S)\}$. This is characterized as the element of Λ_0 which is orthogonal to $\mathcal{N}(S)$. Therefore, $\lambda_0 = \lambda^* + \lambda_N$, where $\lambda_N \in \mathcal{N}(S)$ and under (5)*

$$\phi(\lambda_0) = \phi(\lambda^*).$$

This suggests a general plug-in principle for estimating $\phi(\lambda_0)$, namely $\phi(\lambda^)$. Typically, estimating λ^* is an ill-posed inverse problem, see Carrasco, Florens and Renault (2006). Also, note that under regular identification, by Theorem 3.4, for l_ϕ solving $S^*l_\phi = r_\phi$ and since $S\lambda_0 = 1$,*

$$\phi(\lambda_0) = \langle \lambda_0, r_\phi \rangle_{\mathbf{H}} = \langle \lambda_0, S^*l_\phi \rangle_{\mathbf{H}} = \mathbb{E}[l_\phi(Z)],$$

which suggests a simple analog estimator for $\phi(\lambda_0)$. Establishing conditions for consistency of these generic procedures is beyond the scope of this paper.

Remark 3.4 *Note that if $\mathcal{N}(S^*) \neq \{0\}$ multiple estimating equations can be used for $\phi(\lambda_0)$ in the regular case (as there exist multiple distinct solutions l_ϕ to $S^*l_\phi = r_\phi$). This is consistent with the definition of regular overidentification in Chen and Santos (2015). To see this, note that $\mathcal{R}(S)$ is the linear span of scores of the model and by Theorem 3 in Luenberger (1997, p.157) the tangent space is $\overline{\mathcal{R}(S)} = L_2^0$ iff $\mathcal{N}(S^*) = \{0\}$. Thus, the model is (regularly) overidentified iff $\mathcal{N}(S^*) \neq \{0\}$. There are also important implications of the results of the present paper in connection to and combination with Chen and Santos (2015). For example, by the results below on Information Loss models, just identification, i.e. $\mathcal{N}(S^*) = \{0\}$, can be interpreted as a certain completeness condition. I do not discuss this further, since the focus of this paper is identification, and in particular irregular identification, rather than regular overidentification.*

⁴I note that the results of this paper can be applied mutatis mutandis to conditional moment restrictions models, simply by replacing the score operator by the derivative of a conditional mean operator. A previous working paper version of this paper illustrates this point with an application to a nonparametric asset pricing model.

4 Information Loss Models

This section introduces a relatively large class of models for which score operators take a simple form; they are conditional mean operators after an appropriate reparametrization explained below. Information Loss models are models where the observed data Z is a known measurable transformation of some unobservable variables Z^* (which may include components that are observable), say $Z = m(Z^*)$. Examples 1 and 2 fall under this setting. Another example is the Rubin Causal Model (cf. Rubin, 1974), which is widely used in the social sciences.

Example 5: Rubin Causal Model. The observables are $Z = (Y, D, X)$, where Y is the outcome of interest, $D = 1$ if the individual is treated and 0 otherwise, and X is a vector of covariates. The unobservables are the potential outcomes $Y(1)$ and $Y(0)$, which are related to the outcome through $Y = Y(1)D + Y(0)(1 - D)$. That is, $Z^* = (Y(1), Y(0), D, X)$, $Z = m(Z^*)$ and $m(y(1), y(0), d, x) = (y(1)d + y(0)(1 - d), d, x)$. \blacktriangle

Suppose Z^* has a distribution G_0 with a corresponding density (with respect to a σ -finite measure π) $\lambda_0 \in \Lambda$. The distribution of Z is then $\mathbb{P}_{\lambda_0} = G_0 m^{-1}$. In models where the nonparametric parameter λ_0 is a density, such as the class described here, it is convenient to introduce scores $b \in L_2^0(G_0)$, and re-define the score operator in (1) as

$$Sb \equiv \frac{f_{\lambda_0 + \lambda_0 b} - f_{\lambda_0}}{f_{\lambda_0}} 1(f_{\lambda_0} > 0). \quad (8)$$

Under the mild assumption that $0 < \lambda_0$ π -a.s., such reparameterizations do not have any impact on the identification results above and simplify the interpretation. Similarly, $T(\lambda_0)$ is re-defined as the linear span of $\{b \in L_2^0(G_0) : \lambda_0 + \lambda_0 b \in \Lambda\}$. Often the set of tangents $T(\lambda_0)$ only contains functions that are bounded. This condition implies Assumption 1(iii), as one can find a small $c \equiv c(b) \in \mathbb{R}$ such that $\lambda_0(1 + cb)$ is a density in Λ (integrates up to one because $b \in L_2^0(G_0)$ and $\lambda_0(1 + cb) \geq 0$ π -a.s.).

In a more general setting than the one considered here, Le Cam and Yang (1988) provide conditions under which the score operator for Information Loss models has the form

$$Sb = \mathbb{E}[b(Z^*) | Z = z], \quad (9)$$

where $b \in T(\lambda_0) \subseteq L_2^0(G_0)$ (see e.g. Lemma 25.34 in Van der vaart 1998). Rather than applying the general results of Le Cam and Yang (1988), we will see that (9) holds directly in the examples considered above from the definition in (8).

The conditional mean representation of the score in (9) also implies that for $g \in L_2^0$,

$$S^*g = \Pi_{\overline{T(\lambda_0)}} \mathbb{E}[g(Z) | Z^* = z^*], \quad (10)$$

where henceforth, for a closed subspace V , Π_V denotes its orthogonal projection operator. This conditional mean representation of score operators turns out to be very useful and is exploited in the examples above. When $T(\lambda_0)$ is dense in $L_2^0(G_0)$ the projection in (10) is of course not needed. The simplicity of interpretation and computation of score operators for Information Loss models facilitates the application of the identification conditions found in this paper.

Example 1: Unemployment Duration with Heterogeneity, cont. From (8), the score operator is given by $S : L_2^0(G_0) \rightarrow L_2$

$$\begin{aligned} Sb &= \frac{1}{f_{\lambda_0}(t_1, t_2)} \int f_{z/\alpha, \beta}(t_1, t_2) \lambda_0(\alpha, \beta) b(\alpha, \beta) d\pi(\alpha, \beta) \\ &= \frac{1}{f_{\lambda_0}(t_1, t_2)} \int f_{z, \alpha, \beta}(t_1, t_2) b(\alpha, \beta) d\pi(\alpha, \beta) \\ &= \mathbb{E}[b(\alpha, \beta) | Z = z], \end{aligned}$$

where $z = (t_1, t_2)$ and $f_{z, \alpha, \beta}(t_1, t_2)$ is the joint density of (Z, α, β) , while the adjoint is

$$S^*g = \mathbb{E}[g(Z) | \alpha, \beta], \quad g \in L_2^0.$$

The following result uses this representation of S^* to provide a necessary condition for regular identification of continuous linear functionals in this example.

Proposition 4.1 *In this example*

$$\mathcal{R}(S^*) \subset \left\{ b(\alpha, \beta) \in L_2^0(G_0) : b(\alpha, \beta) = C_1 + C_2 \beta^2 e^{2\alpha\beta} h(\alpha^2, \beta^2) \right\},$$

for constants C_1 and C_2 and a continuous function $h(u, v)$ defined on $(0, \infty)^2$ that, if \mathcal{T} is bounded, is an infinite number of times differentiable at $u \in (0, \infty)$, for all $v \in (0, \infty)$.

Alvarez, Borovicková and Shimer's (2016) nonparametric identification result and Theorem 3 in Luenberger (1997, p.157) imply that $\overline{\mathcal{R}(S^*)} = L_2^0(G_0)$. Thus, Proposition 4.1 shows that the class of irregularly identified functionals is large, i.e. the boundary of $\mathcal{R}(S^*)$ is large. Intuitively, this follows because the density $f_{z/\alpha, \beta}(t_1, t_2)$ is very smooth in the parameters (α, β) , so that $\mathcal{R}(S^*)$ only contains very smooth functions. An implication of Proposition 4.1 is that the cdf of unobserved heterogeneity at the fixed point (α_0, β_0) , i.e. $\phi(\lambda_0) = \mathbb{E}[1(\alpha \leq \alpha_0)1(\beta \leq \beta_0)]$, is not regularly identified. This seems to be a generic feature of many models with nonparametric unobserved heterogeneity, such as that in Heckman and Singer (1984a, 1984b); see further examples below in the context of random coefficient models, and see Van der Vaart (1991) and Bickel et al. (1998, Chapter 6) for analogous results in some exponential and uniform mixture models. \blacktriangle

The representation of the score operator as a conditional mean operator implies the following result, which establishes a link with identification results in the nonparametric instrumental variables literature; see Newey and Powell (2003), Blundell, Chen and Kristensen (2007), Andrews (2011) and D'Haultfoeuille (2011), among others. Its proof is trivial and hence omitted.

Proposition 4.2 *If the representation (9) holds, then $\mathcal{N}(S) = \{0\}$ means that the distribution of Z^* given Z is L_2 -complete on $\overline{T(\lambda_0)}$ (i.e. for all $b \in \overline{T(\lambda_0)} : \mathbb{E}[b(Z^*) | Z] = 0 \implies b = 0$).*

If the parameter space for unobserved heterogeneity is nonparametric, in the sense that $\overline{T(\lambda_0)} = L_2^0$, then, by Proposition 4.2 and Newey and Powell (2003), nonparametric identification imposes limits on the dimensionality and support of unobservables relative to observables. In particular, if Z^* and Z are continuous, then the dimension of Z^* cannot exceed that of Z for nonparametric identification to hold.

Another general implication of the results of this section is that if the support of the observable variables is discrete, then one can only identify a finite number of linear functionals and these functionals can be characterized. I formalize this discussion in the following

Proposition 4.3 *With a score operator (9), if the support of Z is discrete and given by $\{z_1, \dots, z_m\}$, then $\mathcal{R}(S^*)$ is finite-dimensional, then closed, and is generated by the elements*

$$r_j(z^*) = \Pi_{\overline{T(\lambda_0)}} \mathbb{P}[Z = z_j | Z^* = z^*], \text{ for } j = 1, \dots, m. \quad (11)$$

This proposition shows that with discrete supports the only functionals that are identified are those whose Riesz's representer is generated by linear combinations of the $r_j(z^*)$, for $j = 1, \dots, m$, in (11). Identification is always regular with discrete supports, as $\mathcal{R}(S^*)$ is closed. Irregular identification only occurs in models with at least one continuous observable variable.

Example 2: Random Coefficient Binary Choice, cont. In this example the score operator is given by the conditional mean operator

$$Sb = \mathbb{E}[b(\beta_i) | Y_i = 1, X_i = x] = \frac{\int_{\mathbb{S}^{d-1}} 1(x's \geq 0) b(s) \lambda_0(s) d\sigma(s)}{f_{\lambda_0}(x)},$$

whereas the adjoint is given by

$$S^*g = \mathbb{E}[g(Y_i = 1, X_i) | \beta_i = s] = \int 1(x's \geq 0) g(1, x) dv_X(x),$$

where henceforth v_X denotes the probability measure of X .

Suppose that the support of X is discrete, and given by $\{x_1, \dots, x_J\}$. Then, applying Proposition 4.3, $\mathcal{R}(S^*)$ is finite-dimensional and generated by the functions

$$r_j(s) = \mathbb{E}[1(X's \geq 0) 1(X = x_j)], \text{ for } j = 1, \dots, J.$$

These results on identification with discrete regressors are complementary to some partial identification results given for binary choice models with discrete regressors in Magnac and Maurin (2008).

The following result provides a necessary condition for regular identification when X is continuous.

Proposition 4.4 *If the distribution of $X/|X|$ is absolutely continuous, then $\mathcal{R}(S^*)$ consists of uniformly continuous functions on \mathbb{S}^{d-1} . If $X = (1, \tilde{X})$ then for each $b \in \mathcal{R}(S^*)$, $b(s_1, s_2)$ is an absolutely continuous function of s_1 for each s_2 , where s_1 denotes the coefficient associated to the intercept.*

An implication of this proposition is that functionals such as the cdf of random coefficients are not regularly identified. A full characterization of $\mathcal{R}(S)$ under different assumptions on $\overline{T(\lambda_0)}$ is given in Rubin (1999), see also Proposition 3.1 in Gautier and Kitamura (2013). These conditions also characterize $\mathcal{R}(S^*)$, given the symmetry of the “kernel” $1(x's \geq 0)$ in x and s . Details are omitted to save space. For the purpose of establishing irregularity of the cdf and quantiles Proposition 4.4 suffices. I also note that these identification results on random coefficients are complementary to those obtained in e.g. Hoderlein, Nesheim and Simoni (2012), Masten (2015), and references therein. \blacktriangle

5 Semiparametric Models

This section applies the results above to the important class of semiparametric models, where $\mathcal{P} = \{\mathbb{P}_{\theta, \eta} : \theta \in \Theta, \eta \in H\}$. The parameter space $\Lambda = \{(\theta, \eta) : \theta \in \Theta, \eta \in H\}$ is a subset of a Hilbert space $\mathbf{H} = \mathbb{R}^p \times \mathcal{H}$. Define $\langle (\theta_1, \eta_1), (\theta_2, \eta_2) \rangle_{\mathbf{H}} := \theta_1' \theta_2 + \langle \eta_1, \eta_2 \rangle_{\mathcal{H}}$. For semiparametric models the score operator has the representation

$$S(b_\theta, b_\eta) = \dot{l}'_\theta b_\theta + \dot{l}_\eta b_\eta, \quad b = (b_\theta, b_\eta) \in \mathbf{H}, \quad (12)$$

where $\dot{l}_\theta \in L_2^p$ is the ordinary score function of θ and \dot{l}_η is a continuous linear operator from $T(\eta_0) \subset \mathcal{H}$ to L_2 . Let $\tilde{l}_\theta := \dot{l}_\theta - \Pi_{\overline{\mathcal{R}(\dot{l}_\eta)}} \dot{l}_\theta$ be the so-called efficient score function for θ . The efficient Fisher information matrix for θ is then given by $\tilde{I}_\theta := \mathbb{E} [\tilde{l}_\theta \tilde{l}_\theta']$. The following result provides a representation of the main identification condition for the finite-dimensional parameter of a semiparametric model. For simplicity, I consider the case $p = 1$, the extension to $p > 1$ is straightforward.

Proposition 5.1 *For the functional $\phi(\lambda) = \theta \in \mathbb{R} : \mathcal{N}(S) \subset \mathcal{N}(\dot{\phi})$ holds iff (i) $\dot{l}_\theta \notin \overline{\mathcal{R}(\dot{l}_\eta)}$ (positive information $\tilde{I}_\theta > 0$) or (ii) $\dot{l}_\theta \in \overline{\mathcal{R}(\dot{l}_\eta)} \setminus \mathcal{R}(\dot{l}_\eta)$ (zero information $\tilde{I}_\theta = 0$).*

Proposition 5.1 shows that identification for the finite-dimensional parameter θ depends crucially on the position of the score \dot{l}_θ of the parameter of interest relative to the linear span of scores for nuisance parameters, i.e. $\mathcal{R}(\dot{l}_\eta)$, and its mean square closure $\overline{\mathcal{R}(\dot{l}_\eta)}$.

Remark 5.1 *The corresponding r_ϕ to $\phi(\lambda) = \theta$ is $r_\phi = (1, 0) \in \mathbb{R} \times \mathcal{H}$. Then, by Proposition 3.3 $\mathcal{N}(S) \subset \mathcal{N}(\dot{\phi})$ implies $0 \in \overline{\mathcal{R}(\dot{l}_\eta^*)}$. If $\mathcal{R}(\dot{l}_\eta^*)$ is closed, the last condition becomes $0 = \dot{l}_\eta^* g$ for some $g \neq 0$. Chamberlain (2010) and Johnson (2004) have shown that in some discrete choice models $0 = \dot{l}_\eta^* g$, for some $g \neq 0$, is necessary for identification. See also Buchinsky, Hahn and Kim (2010) for an application of this result in structural discrete choice models.*

Remark 5.2 *If \dot{l}_η^* satisfies a completeness condition, in the sense that $\mathcal{N}(\dot{l}_\eta^*) = 0$, then $\overline{\mathcal{R}(\dot{l}_\eta)} = L_2^0$, by Theorem 3 in Luenberger (1997, p.157), and as a result θ_0 has zero information, since $\tilde{l}_\theta = \dot{l}_\theta - \dot{l}_\theta = 0$ and hence $\tilde{I}_\theta = 0$. This general structure underlies in the zero information calculations of Chamberlain (2010) with nonlinear panel data binary choice with fixed effects, Ishwaran (1999) for some semiparametric mixtures of exponential families, Bajari, Hahn, Hong and Ridder (2011) for games*

with multiple equilibria, and the zero information for the non-random coefficients in the semiparametric mixed Logit given below, among others.

Similarly, we may be interested in functionals of the nuisance parameter, $\phi(\lambda) = \chi(\eta)$, where $\chi : \mathcal{H} \mapsto \mathbb{R}$ is a linear continuous functional. Let $r_\chi \in T(\eta_0) \subset \mathcal{H}$ such that for all $b_\eta \in T(\eta_0)$

$$\chi(b_\eta) = \langle b_\eta, r_\chi \rangle_{\mathcal{H}}.$$

Keep for simplicity the assumption that $\theta \in \mathbb{R}$; this can be relaxed at the cost of introducing further notation. I then obtain a similar characterization to that of Proposition 5.1, allowing for singular information for both the parametric part θ and the functional $\phi(\lambda) = \chi(\eta)$. Let b_η^0 be the minimum norm solution of the equation $\dot{l}_\theta = \dot{l}_\eta b_\eta$ given by (provided it exists)

$$b_\eta^0 = \left(\dot{l}_\eta^* \dot{l}_\eta \right)^- \dot{l}_\eta^* \dot{l}_\theta,$$

where A^- denotes the generalized inverse of A .

Proposition 5.2 *For the functional $\phi(\lambda) = \chi(\eta) \in \mathbb{R}$: (i) if $\dot{l}_\theta \notin \mathcal{R}(\dot{l}_\eta)$, then $\mathcal{N}(S) \subset \mathcal{N}(\dot{\phi})$ holds iff $r_\chi \in \overline{\mathcal{R}(\dot{l}_\eta^*)}$; (ii) if $\dot{l}_\theta \in \mathcal{R}(\dot{l}_\eta)$, then $\mathcal{N}(S) \subset \mathcal{N}(\dot{\phi})$ holds iff $r_\chi \in \overline{\mathcal{R}(\dot{l}_\eta^*)}$ and $\langle b_\eta^0, r_\chi \rangle_{\mathcal{H}} = 0$.*

Remark 5.3 *The conditions for identification of $\chi(\eta_0)$ depend on whether θ_0 is identified or not. In both cases $r_\chi \in \overline{\mathcal{R}(\dot{l}_\eta^*)} \setminus \mathcal{R}(\dot{l}_\eta^*)$ corresponds to the case of zero information for $\phi(\lambda) = \chi(\eta)$ at $\phi(\lambda_0) = \chi(\eta_0)$. Regular identification of $\chi(\eta)$ in case (ii) requires that for all r_χ^* that solve $r_\chi = \dot{l}_\eta^* r_\chi^*$ it holds that $\langle \dot{l}_\theta, r_\chi^* \rangle = 0$ (which can be shown to imply $\langle b_\eta^0, r_\chi \rangle_{\mathcal{H}} = 0$). Under this condition, lack of identification of θ_0 does not affect regular identification of $\chi(\eta_0)$.*

In the remainder of this section I extend some of the previous results to a class of models that are nonlinear in the parameter of interest but linear in nuisance parameters. For simplicity, these results are given for the finite-dimensional parameter θ , but analogous results can be obtained for a functional $\phi(\lambda) = \chi(\eta)$, see Section 8.4 in the Appendix. Assume the score operator in (12) is linear and consider models satisfying

Assumption 2: For some $\rho \geq 1$ and for all $\varepsilon > 0$, there exists $\delta > 0$ such that,

$$\|(f_\lambda - f_{\lambda_0}) / f_{\lambda_0} - S(\lambda - \lambda_0)\| < \varepsilon |\theta - \theta_0|^\rho, \text{ for all } \lambda = (\theta, \eta) \in \mathcal{B}_\delta(\lambda_0).$$

Assumption A2 is a mean-square differentiability condition with a Lipschitz property on the derivative. It generally holds for models that are nonlinear and smooth in the parameter of interest θ , but linear in the nuisance parameters. Examples of models in this class include, among others, linear and nonlinear panel data models with fixed effects (see Example 3 above); incomplete and complete games with multiple equilibria (see e.g. Bajari, Hahn, Hong and Ridder 2011); semiparametric measurement error models (see e.g. Hu and Schennach 2008); dynamic discrete choice models (see e.g. Hu and Shum

2012); and binary discrete choice models with single and multiple agents (see e.g. Chamberlain 1986, and more recently, Khan and Nekipelov 2016).

To accommodate cases with zero information, i.e. irregular identification, I define a “generalized Fisher information”

$$I_{\theta,\rho} = \inf_{\substack{b \in \overline{T(\lambda_0)} \\ b = (\theta - \theta_0, b_\eta), \theta \neq \theta_0}} \frac{\|Sb\|^2}{|\theta - \theta_0|^{2\rho}}, \quad (13)$$

where $1 \leq \rho < \infty$. It is straightforward to show that $\tilde{I}_\theta \equiv I_{\theta,1}$. That is, the classical semiparametric Fisher information corresponds to $\rho = 1$. Moreover, it can be shown that $\tilde{I}_\theta \equiv I_{\theta,1} \leq I_{\theta,\rho}$ for $1 < \rho < \infty$ (this follows because in the definition of $I_{\theta,\rho}$ we can assume $|\theta - \theta_0| \leq 1$ without loss of generality).

Theorem 5.3 *Let Assumption 2 hold. Then, $I_{\theta,\rho} > 0$ implies local identification of θ at θ_0 .*

Corollary 5.1 *Let Assumption 2 hold. Then, the parameter θ_0 is locally regularly identified if $\tilde{I}_\theta > 0$ (which corresponds to $\dot{l}_\theta \notin \overline{\mathcal{R}(\dot{l}_\eta)}$) and locally irregularly identified if $\tilde{I}_\theta = 0$ but $I_{\theta,\rho} > 0$ for $1 < \rho < \infty$ (which corresponds to $\dot{l}_\theta \in \mathcal{R}(\dot{l}_\eta) \setminus \overline{\mathcal{R}(\dot{l}_\eta)}$).*

Theorem 5.3 extends the sufficient condition of Rothenberg (1971) to the semiparametric, possibly irregular, case analyzed here. Theorem 7 in Chen et al. (2014) gives related semiparametric local identification results for the regular case, $\rho = 1$, in conditional moment models. Chamberlain (1992) discusses conditions for $\tilde{I}_\theta > 0$ in conditional moment models. Section 8.2 in the Appendix provides sufficient and necessary conditions for a positive generalized Fisher information $I_{\theta,\rho} > 0$, with $1 \leq \rho < \infty$, in terms of the so called Singular Value Decomposition of the information operator and “smoothness” of the Riesz representer r_ϕ . For many examples where the score operator is a “smoothing” operator, one can index the level of irregularity in the identification of $\phi(\lambda_0)$ by the smoothness of the function r_ϕ (see Kress 1999, Chapter 8), with a smoother r_ϕ corresponding to more regularity.

For simplicity of presentation, I have focussed in Theorem 5.3 on what is often called “mild irregularity”, but note that these results can be extended to general degrees of irregularity simply by replacing $|\theta - \theta_0|^\rho$ in Assumption 2 and in the definition of the generalized information by $\psi(|\theta - \theta_0|)$, for a function ψ that is non-decreasing, non-negative, which is right continuous at 0 and with $\psi(0) = 0$, see Section 8.3 for an extension along these lines and a set of alternative regularity conditions. With these changes different degrees of irregularity are allowed. The case $\psi(\epsilon) = \epsilon^\rho$, $1 < \rho < \infty$, corresponds to mild or moderate irregularity, while $\psi(\epsilon) = \exp(\epsilon) - 1$, for example, corresponds to severe irregularity. In applications such as for example those in structural models of unemployment (e.g. Heckman and Singer 1984a, 1984b, Alvarez et al. 2016), nonparametric instrumental variables and the semiparametric mixed Logit model, it is imperative to allow for severe irregularity in the sufficient conditions for identification.

6 Examples

6.1 Random Coefficient Logit Model

The random coefficients Logit model, also known as the mixed Logit, is one of the most commonly used models in applied choice analysis. It was introduced by Boyd and Mellman (1980) and Cardell and Dunbar (1980) and it is widely used in environmental economics, industrial economics, marketing, public economics, transportation economics and other fields. Fox, Kim, Ryan and Bajari (2012) have recently shown that the semiparametric mixed Logit model is nonparametrically identified. Here, it is shown that identification of the non-random coefficients θ_0 is necessarily irregular. As a result, there does not exist a regular estimator of θ_0 converging at the parametric rate and the root-n estimability often used in the applied literature is shown to necessarily depend on functional form assumptions for the distribution of the random coefficients.

The data $Z_i = (Y_i, X_i)$ is a random sample from the density (with respect to μ below),

$$f_{\lambda_0}(y, x) = \int f_{y/x, \beta}(y; \theta_0) \eta_0(\beta) d\beta,$$

where $\lambda_0 = (\theta_0, \eta_0) \in \Theta \times H$, $\theta_0 = (\theta_{01}, \dots, \theta_{0J})'$,

$$f_{y/x, \beta}(y; \theta_0) = \frac{\exp(\theta_{0y} + x'_y \beta)}{1 + \sum_{j=1}^J \exp(\theta_{0j} + x'_j \beta)},$$

for $x = (x_0, x_1, \dots, x_J) \in \mathcal{X}$ and $y \in \mathcal{Y} = \{0, 1, \dots, J\}$. The consumer can choose between $j = 1, \dots, J$, $J < \infty$, mutually exclusive inside goods and one outside good ($y = 0$). The probability for the inside good is normalized so that $\theta_{00} = 0$ and $x_0 = 0$. The random coefficients have a Lebesgue density $\eta_0(\beta)$ and are independent of covariates. More general versions of the model that allow for coefficients θ_{0y} depending on other covariates and/or endogenous regressors lead to the same conclusions regarding irregular identifiability of fixed parameters. Thus, we keep the simple version here. The measure μ is defined on $\mathcal{Z} = \mathcal{Y} \times \mathcal{X}$ as $\mu(B_1 \times B_2) = \tau(B_1) \nu_X(B_2)$, where $B_1 \subset \mathcal{Y}$, B_2 is a Borel set of \mathcal{X} , $\tau(\cdot)$ is the counting measure and $\nu_X(\cdot)$ is the probability measure for X . The vector β and covariates x_y are K -dimensional. The parameter space Θ is an open set of \mathbb{R}^J . The set H consists of measurable functions $\eta : \mathbb{R}^K \rightarrow \mathbb{R}$ such that (i) $\inf_{\beta \in B} \eta(\beta) > 0$ for any compact set B of \mathbb{R}^K ; and (ii) $\int \eta(\beta) d\beta = 1$. Let x_{-j} denote all components of x but x_j . The following assumption is needed.

Assumption 3: For each $j = 1, \dots, J$: x_j contains zero as an element of the interior of its support, $\mathbb{E}[x_j x'_j]$ is positive definite, and x_j is variation free from x_{-j} .

These conditions are assumed in Fox et al. (2012), and are used and extensively discussed there, together with other conditions on the distribution of X , to nonparametrically identify $\lambda_0 = (\theta_0, \eta_0)$.

Theorem 6.1 *Let Assumption 3 hold. Then, the semiparametric Fisher information for θ_{0j} in the random coefficients Logit is zero for all $j = 1, \dots, J$.*

The proof of Theorem 6.1 follows closely that of Chamberlain (2010, Theorem 2) for binary choice panel data. I make explicit in the proof that the completeness condition $\mathcal{R}(\dot{i}_\eta) = L_2^0$ holds, which implies the zero information for θ_{0j} (see Remark 5.2). For $J = 1$, a simpler method of proof checks directly that $\mathcal{N}(\dot{i}_\eta^*) = 0$, which follows by well known results in consistent specification testing (Stinchcombe and White, 1998). The completeness condition here is also related to the general approximation theorem provided in McFadden and Train (2000, Theorem 1) for mixed logit models. McFadden and Train consider different conditions on covariates and do not investigate identification.

Despite the zero information for θ_0 , as shown by Fox et al. (2012) still this parameter is identified. Indeed, under Assumption 3

$$\theta_{0j} = \log f_{\lambda_0}(y = j, x = 0) - \log f_{\lambda_0}(y = 0, x = 0), \quad j = 1, \dots, m.$$

Identification of θ_0 comes from a thin set of measure zero (x at zero). Our results imply that this is not specific to this way of identifying θ_0 , but rather that any possible identification result of θ_0 in this model must have a similar feature, and that as a result there does not exist a regular estimator of θ_0 converging at the parametric rate.

The results above on functionals of the random coefficients' distribution can also be used here to investigate which functionals of unobserved heterogeneity are regularly or irregularly identified. Proposition 5.2 and the remark that follows show that regular functionals $\phi(\lambda) = \chi(\eta) \in \mathbb{R}$ are those with a Riesz representer $r_\chi \in \mathcal{R}(\dot{i}_\eta^*)$. In the random coefficients Logit, $\mathcal{R}(\dot{i}_\eta^*)$ is a set of analytic functions (functions that are infinitely differentiable with a convergent power series expansion). Thus, the class of regular functionals is relatively small, and functionals such as the cumulative distribution function, quantile, or other functionals with discontinuous influence functions are only irregularly identified.

6.2 Nonlinear Fixed Effects Panel Data: Binary Choice

As an illustration, I consider Chamberlain's (2010) static binary choice setting. Note, however, that the theory above applies to other nonlinear panel data models where the conditional likelihood given unobserved heterogeneity is parametric. The binary choice case has a long history in econometrics, and its identification is challenging. Here, I revisit Chamberlain's (2010) identification analysis of the parametric component using the tools of this paper and I also discuss identification of AME.

The researcher observes the vector $Z_i = (Y_{i1}, Y_{i2}, X_{i1}, X_{i2})$, i.e. $T = 2$, of a binary dependent variable satisfying

$$Y_{it} = 1 \left(\alpha_i + \theta_{01}d_{it} + \theta'_{02}X_{it} \geq u_{it} \right), \quad (14)$$

where u_{it} are *iid* variables with cdf F , independent of $X_i = (X_{i1}, X_{i2})$, $d_{it} = 1$ if $t = 2$ and zero otherwise, and $\theta_0 = (\theta_{01}, \theta'_{02})'$. The support of X_i is $\mathcal{X} \subseteq \mathbb{R}^{2p}$. The conditional density $f_{y/x, \alpha}$ is given by

$$f_{y/x, \alpha}(y; \theta_0) = F(\alpha + \theta'_{02}x_1)^{y_1} (1 - F(\alpha + \theta'_{02}x_1))^{1-y_1} F(\alpha + \theta_{01} + \theta'_{02}x_2)^{y_2} (1 - F(\alpha + \theta_{01} + \theta'_{02}x_2))^{1-y_2},$$

where $x = (x_1, x_2) \in \mathcal{X}$ and $y = (y_1, y_2) \in \mathcal{Y} = \{(1, 1), (1, 0), (0, 1), (0, 0)\}$. Let $\eta_0(\alpha, x)$ denote the unrestricted conditional (Lebesgue) density of α given $X = x$. The measure μ is defined on $\mathcal{Z} = \mathcal{Y} \times \mathcal{X}$

as $\mu(B_1 \times B_2) = \tau(B_1)\nu_X(B_2)$, where $B_1 \subset \mathcal{Y}$, B_2 is a Borel set of \mathcal{X} , $\tau(\cdot)$ is the counting measure and $\nu_X(\cdot)$ is the probability measure for X . Then, the density of the observation $Z_i = (Y_i, X_i)$ with respect to μ is

$$f_{\lambda_0}(y, x) = \int f_{y/x, \alpha}(y; \theta_0) \eta_0(\alpha, x) d\alpha,$$

where $\lambda_0 = (\theta_0, \eta_0)$. Under the conditions spelled in Chamberlain (2010) the score operator is given by

$$S(b_\theta, b_\eta) = \dot{l}'_\theta b_\theta + \dot{l}'_\eta b_\eta,$$

where

$$\dot{l}_\theta = \frac{1}{f_{\lambda_0}(y, x)} \int \frac{\partial f_{y/x, \alpha}(y; \theta_0)}{\partial \theta} \eta_0(\alpha, x) d\alpha$$

and

$$\begin{aligned} \dot{l}_\eta b_\eta &= \frac{1}{f_{\lambda_0}(y, x)} \int f_{y/x, \alpha}(y; \theta_0) b_\eta(\alpha, x) \eta_0(\alpha, x) d\alpha \\ &= \mathbb{E}[b_\eta(\alpha, x) | Y = y, X = x], \end{aligned}$$

where $b_\theta \in \mathbb{R}^p$, $b_\eta \in L_2^0(G_0)$ and G_0 is the joint distribution of (α_i, X_i) .

I start with an analysis of the identification of $\phi(\theta, \eta) = \theta$. Note that

$$\begin{aligned} \dot{l}_\eta^* g &= \mathbb{E}[g(Y_i, X_i) | \alpha_i = \alpha, X_i = x] \\ &= \sum_{j=1}^4 g_j(x) p_j(\alpha, x), \end{aligned}$$

where $g_j(x) = g(\tilde{y}_j, x)$, $p_j(\alpha, x) \equiv p_{j, \theta_0}(\alpha, x)$, with $p_{j, \theta}(\alpha, x) = \mathbb{P}_{\theta, \eta_0}[Y_i = \tilde{y}_j | \alpha_i = \alpha, X_i = x]$, and $\tilde{y}_j = (1, 1), (1, 0), (0, 1)$ and $(0, 0)$, for $j = 1, \dots, 4$, respectively. Therefore

$$\mathcal{N}(\dot{l}_\eta^*) = \left\{ g \in L_2^0 : \sum_{j=1}^4 g_j(x) p_j(\alpha, x) = 0 \right\}.$$

Chamberlain (2010) showed that if F is not logistic then $\mathcal{N}(\dot{l}_\eta^*) = \{0\}$, which implies the zero information for θ by the Remark 5.2 above. Indeed, Chamberlain (2010) computed the information \tilde{I}_θ and showed that it was zero. The zero information follows because the linear span of scores for nuisance parameters can approximate arbitrarily well any zero mean function with finite variance, and in particular \dot{l}_θ . This approach of checking for completeness, $\mathcal{N}(\dot{l}_\eta^*) = \{0\}$, is relatively simple here because Y is discrete.

The implication of Chamberlain's (2010) result is that regular identification of θ_0 in (14) with fixed effects is not possible when F is not Logistic. The results of the present paper can be used to assess the sensitivity of regular identification of θ_0 to different prior assumptions on unobserved heterogeneity for other cases, such as Probit. By Remark 5.2 and the results above a necessary condition for regular identification of θ_0 with a generic parameter space for heterogeneity is

$$\sum_{j=1}^4 g_j(x) \Pi_{\overline{T(\eta_0)}} p_j(\alpha, x) = 0 \text{ for a.s. } \alpha, x, \quad (15)$$

for some non-zero g 's. This can be the basis for a systematic study of regular identification in binary choice panel data models. Such investigation will be carried out elsewhere.

Suppose now that one is interested in the identification of the AME

$$\phi(\lambda_0) = \mathbb{E}[m(\alpha_i, X_i)],$$

for some $m \in L_2(G_0)$. Then, by Proposition 5.2 the assumption to check is $m \in \overline{\mathcal{R}(\dot{l}_\eta^*)}$ (note $r_\chi = m$ in this example). Regular identification requires $m \in \mathcal{R}(\dot{l}_\eta^*)$, which in this model means, for some $g \in L_2^0$,

$$\dot{l}_\eta^* g = \sum_{j=1}^4 g_j(x) p_j(\alpha, x) = m(\alpha, x). \quad (16)$$

Thus, the number of marginal effects that are regularly identified is rather small, in the sense that $m(\cdot, x)$ is generated by the four dimensional space spanned by $p_j(\cdot, x)$, for $j = 1, \dots, 4$. There are two cases: (i) bounded support of X ; and (ii) unbounded support of X . For (ii), regular identification holds if m satisfies (16). In case (i), θ_0 is not identified as shown by Chamberlain (2010), however it is still possible that the AME $\phi(\lambda_0) = \mathbb{E}[m(\alpha_i, X_i)]$ is regularly identified. By Proposition 5.2(ii) this is possible if in addition to (16), m is orthogonal to the minimum norm solution of $\dot{l}_\eta b_\eta = \dot{l}_\theta$. Using the definitions of \dot{l}_η and \dot{l}_η^* this holds if for any $b_\eta(\cdot)$ satisfying $\dot{l}_\eta b_\eta = \dot{l}_\theta$ and any m satisfying (16),

$$\begin{aligned} \mathbb{E}[m(\alpha_i, X_i) b_\eta(\alpha_i, X_i)] &= \mathbb{E}[\mathbb{E}[g(Y_i, X_i) | \alpha_i, X_i] b_\eta(\alpha_i, X_i)] \\ &= \mathbb{E}[g(Y_i, X_i) b_\eta(\alpha_i, X_i)] \\ &= \mathbb{E}[g(Y_i, X_i) \mathbb{E}[b_\eta(\alpha_i, X_i) | Y_i, X_i]] \\ &= \mathbb{E}[g(Y_i, X_i) \dot{l}_\theta(Y_i, X_i)] \\ &= 0. \end{aligned}$$

Thus, regular identification of the AME when θ_0 is not identified requires that m is of the form in (16) with the g 's satisfying the orthogonality conditions $\mathbb{E}[g(Y_i, X_i)] = 0$ and $\mathbb{E}[g(Y_i, X_i) \dot{l}_\theta(Y_i, X_i)] = 0$. I summarize the discussion above in the following

Proposition 6.2 *For the binary panel data model in (14), under Chamberlain's (2010) conditions the set of regularly locally identified AME $\phi(\lambda_0) = \mathbb{E}[m(\alpha_i, X_i)]$ are those where m satisfies (16) for $g \in L_2^0$ when θ_0 is identified. If θ_0 is not identified and $\dot{l}_\theta \in \mathcal{R}(\dot{l}_\eta)$, then regular identification of $\phi(\lambda_0)$ follows if in addition $\mathbb{E}[g(Y_i, X_i) \dot{l}_\theta(Y_i, X_i)] = 0$ for all $g \in L_2^0$ satisfying (16).*

The results given here on identification are complementary to the partial identification results obtained in Honoré and Tamer (2006), Chernozhukov et al. (2013) and the irregular identification results of Manski (1987), Honoré and Kyriazidou (2000), Hahn (2001), Bonhomme (2011) and Graham and Powell (2012). What appears to be new here is the connection with completeness; the necessary condition (15) for regular identification with a generic parameter space for heterogeneity; and the identification results for AME when θ_0 is unknown but identified or possibly unidentified.

7 Conclusions

This paper investigates semiparametric identification, with a particular emphasis on irregular identification. The question of semiparametric identification is studied first in a setting where densities are affine in a nonparametric parameter. In this setting, necessary and sufficient conditions for regular and irregular identification, respectively, are obtained. The characterization of irregular identification is used to show that semiparametric irregular identification is a common feature of many economic models of practical interest. For example, structural models with densities that are smooth in parameters indexing the nonparametric unobserved heterogeneity are canonical examples of models with many irregularly identified functionals. The example application in Alvarez, Borovicková and Shimer (2016) illustrates this general point. Functionals of the distribution of unobserved heterogeneity such as the cumulative distribution function, quantiles or other functionals with discontinuous influence functions are generally only irregularly identified in many models that involve nonparametric unobserved heterogeneity. Other irregular identification results are proved in models satisfying certain completeness conditions, as illustrated with the non-random coefficients of the semiparametric mixed Logit model.

The question of whether zero information corresponds to a lack of identification is a rather delicate question, as pointed out in Chamberlain (1986). It is shown here that the position of the score of the parameter of interest relative to the linear span of scores for nuisance parameters and its mean-square closure is the factor determining the regularity of identification, with irregularity corresponding to the boundary case. It is also shown that under mild smoothness conditions a positive semiparametric Fisher information for the parameter implies its local identification. When the semiparametric Fisher information is zero, positivity of a new generalized Fisher information quantity introduced in this paper is shown to imply irregular identification.

There are a number of important extensions of the present study. I provide in Section 8.4 in the Appendix sufficient conditions for regular and irregular semiparametric identification in nonlinear models and for nonlinear functionals. Applying these results to specific examples and establishing connections with rates of convergence for semiparametric estimators remain topics for future research.

8 Appendix

I will extensively use basic results from operator theory and Hilbert spaces in this Appendix. See Carrasco, Florens and Renault (2006) for an excellent review of these results. This Appendix is organized as follows. Section 8.1 provides mathematical proofs of the main results. Section 8.2 discusses further results on regular and irregular identification in terms of the Singular Value Decomposition of the information operator. Section 8.3 establishes sufficient conditions for irregular identification in linear models where the completeness condition does not hold. Finally, Section 8.4 establishes sufficient conditions for identification in general nonlinear models.

8.1 Proofs

Proof of Theorem 3.1: That (3) is a sufficient condition for identification of $\phi(\lambda_0)$ follows from the definition of identification. If $f_\lambda = f_{\lambda_0}$, then $b = \lambda - \lambda_0 \in \mathcal{N}(S) \subset \mathcal{N}(\dot{\phi})$, and hence $\phi(\lambda_0 + b) = \phi(\lambda_0)$. To prove the necessity, suppose that (3) does not hold, i.e.

$$\mathcal{N}(S) \subsetneq \mathcal{N}(\dot{\phi}),$$

then there exists $b \in T(\lambda_0)$ such that $b \in \mathcal{N}(S)$ but $b \notin \mathcal{N}(\dot{\phi})$. This means by linearity that for all $c \in \mathbb{R}$, $cb \in \mathcal{N}(S)$ but $cb \notin \mathcal{N}(\dot{\phi})$. By Assumption 1(iii) there exists c such that $\lambda_0 + cb \in \Lambda$, $cb \in \mathcal{N}(S)$ and $cb \notin \mathcal{N}(\dot{\phi})$. This implies that $f_{\lambda_0+cb} = f_{\lambda_0}$ but $\phi(\lambda_0 + cb) \neq \phi(\lambda_0)$. That is, $\phi(\lambda_0)$ is not identified. ■

Proof of Proposition 3.2: Define $L : L_1(\pi) \mapsto L_1(\mathbb{P})$ as

$$Lb = \int f_{z/\alpha,\beta}(t_1, t_2) b(\alpha, \beta) d\pi(\alpha, \beta).$$

Decompose L as

$$\begin{aligned} Lb &= \int_{\alpha \geq 0} f_{z/\alpha,\beta}(t_1, t_2) b(\alpha, \beta) d\pi(\alpha, \beta) + \int_{\alpha \leq 0} f_{z/\alpha,\beta}(t_1, t_2) b(\alpha, \beta) d\pi(\alpha, \beta) \\ &\equiv L_+ b + L_- b. \end{aligned}$$

Theorem 1 in Alvarez et al. (2016) shows that both L_+ and L_- are linear injective operators, and therefore have inverses, L_+^{-1} and L_-^{-1} , respectively. Define the normalizing positive constant

$$C_L = \int f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) \lambda_0(\alpha, \beta) d\pi(\alpha, \beta) dt_1 dt_2.$$

Then, using that the inverse Gaussian satisfies

$$f(t; \alpha, \beta) = e^{2\alpha\beta} f(t; -\alpha, \beta),$$

it can be shown that

$$\begin{aligned} L_- b &= C_L^{-1} \int_{\alpha \leq 0} e^{4\alpha\beta} f(t_1; -\alpha, \beta) f(t_2; -\alpha, \beta) b(\alpha, \beta) d\pi(\alpha, \beta) \\ &= -C_L^{-1} \int_{\alpha \geq 0} e^{-4\alpha\beta} f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) b(-\alpha, \beta) d\pi(\alpha, \beta) \\ &= -L_+(e^{-4\alpha\beta} b(-\alpha, \beta)). \end{aligned}$$

Then, using these results, $b \in \mathcal{N}(L)$, i.e. $L_+b + L_-b = 0$, is equivalent to

$$\begin{aligned} b(\alpha, \beta) &= -L_+^{-1}(L_-b) \\ &= L_+^{-1}L_+(e^{-4\alpha\beta}b(-\alpha, \beta)) \\ &= e^{-4\alpha\beta}b(-\alpha, \beta). \end{aligned}$$

This concludes the proof after noticing that $\mathcal{N}(S) = \mathcal{N}(L)$. ■

Proof of Proposition 3.3: By Proposition 1 in Luenberger (1997, p.52)

$$\mathcal{N}(S) \subset \mathcal{N}(\dot{\phi})$$

is equivalent to

$$\mathcal{N}(\dot{\phi})^\perp \subset \mathcal{N}(S)^\perp,$$

since both $\mathcal{N}(S)$ and $\mathcal{N}(\dot{\phi})$ are closed linear subspaces, and where henceforth V^\perp denotes the orthocomplement of the subspace V . However, since

$$\dot{\phi}(b) = \langle b, r_\phi \rangle_{\mathbf{H}},$$

for all $b \in T(\lambda_0)$, it follows that $\mathcal{N}(\dot{\phi})^\perp = \text{span}\{r_\phi\}$. On the other hand, by Theorem 3 in Luenberger (1997, p.157)

$$\mathcal{N}(S)^\perp = \overline{\mathcal{R}(S^*)},$$

which proves the Proposition. ■

Proof of Theorem 3.4: By Proposition 3.3, (3) is equivalent to (6). Therefore, the necessity and sufficiency of identification follows from Proposition 3.3 and Theorem 3.1. The qualification of regular or irregular follows from Theorem 4.1 in van der Vaart (1991), which shows that $r_\phi \in \mathcal{R}(S^*) \iff I_\phi > 0$. ■

Proof of Proposition 3.5: Part (i) follows from $\mathcal{N}(I_{\lambda_0}) = \mathcal{N}(S)$ and Theorem 3.1. If I_{λ_0} is regularly positive, then $\mathcal{N}(I_{\lambda_0}) = \{0\}$, and by Theorem 3.1 nonparametric identification holds. For the reciprocal, nonparametric identification implies by Proposition 3.3 that $\overline{\mathcal{R}(S^*)} = \overline{T(\lambda_0)}$. By Cauchy-Schwarz's inequality and $\mathcal{R}(S^*) = \overline{\mathcal{R}(S^*)}$

$$\begin{aligned} \sup_{b \in \overline{\mathcal{R}(S^*)}} \frac{\|b\|_{\mathbf{H}}^2}{\|Sb\|^2} &= \sup_{\substack{b=S^*b^* \\ \|b^*\| \leq 1}} \frac{\|b\|_{\mathbf{H}}^2}{\|Sb\|^2} \\ &= \sup_{\substack{b=S^*b^* \\ \|b^*\| \leq 1}} \frac{\langle b, S^*b^* \rangle_{\mathbf{H}}}{\|Sb\|^2} \\ &= \sup_{\substack{b=S^*b^* \\ \|b^*\| \leq 1}} \frac{\langle Sb, b^* \rangle}{\|Sb\|^2} \\ &\leq 1. \end{aligned} \tag{17}$$

■

Proof of Proposition 4.1: By substitution of $f_{z/\alpha,\beta}(t_1, t_2)$ we obtain

$$\begin{aligned} S^*g &= \mathbb{E}[g(Z)|\alpha, \beta]. \\ &= \int_{\mathcal{T}^2} g(t_1, t_2) f_{z/\alpha,\beta}(t_1, t_2) dt_1 dt_2 \\ &= C\beta^2 e^{2\alpha\beta} h(\alpha^2, \beta^2), \end{aligned}$$

where

$$h(u, v) = \int_{\mathcal{T}^2} g(t_1, t_2) \frac{1}{t_1^{3/2} t_2^{3/2}} s(u, v; t_1) s(u, v; t_2) dt_1 dt_2$$

and

$$s(u, v; t) = \exp\left(-\frac{ut}{2} - \frac{v}{2t}\right), \quad t \in \mathcal{T}, \quad (u, v) \in (0, \infty).$$

We check that the conditions for an application of the Leibniz's rule hold. These conditions are

1. The partial derivative $\partial^m s(u, v; t_1) s(u, v; t_2) / \partial^m u$ exists and is a continuous function on an open neighborhood B of (u, v) , for a.s. $(t_1, t_2) \in \mathcal{T}^2$.
2. There is a positive function $h_m(t_1, t_2)$ such that

$$\sup_{(u,v) \in B} \left| \frac{\partial^m s(u, v; t_1) s(u, v; t_2)}{\partial^m u} \right| \leq h_m(t_1, t_2) \quad (18)$$

and

$$\int_{\mathcal{T}^2} g(t_1, t_2) \frac{1}{t_1^{3/2} t_2^{3/2}} h_m(t_1, t_2) dt_1 dt_2 < \infty. \quad (19)$$

Simple differentiation and induction show that for any integer $m \geq 0$

$$\frac{\partial^m s(u, v; t_1) s(u, v; t_2)}{\partial^m u} = 2^{-m} (-1)^m (t_1 + t_2)^m s(u, v; t_1) s(u, v; t_2).$$

Therefore, by monotonicity we can find u^* and v^* such that (18) holds with

$$h_m(t_1, t_2) = 2^{-m} (t_1 + t_2)^m s(u^*, v^*; t_1) s(u^*, v^*; t_2).$$

Furthermore, by $\mathbb{E}[g(Z)|\alpha, \beta] < \infty$ for all α and β , and the boundedness of \mathcal{T} , condition (19) holds. The continuity of $h(u, v)$ is proved using similar arguments (i.e. for $m = 0$). Details are omitted to save space. ■

Proof of Proposition 4.3: Write for $g \in L_2^0$

$$\begin{aligned} S^*g &= \Pi_{\overline{T(\lambda_0)}} \mathbb{E}[g(Z)|Z^* = z^*]. \\ &= \Pi_{\overline{T(\lambda_0)}} \sum_{j=1}^m g(z_j) \mathbb{P}[Z = z_j | Z^* = z^*] \\ &= \sum_{j=1}^m g(z_j) \Pi_{\overline{T(\lambda_0)}} \mathbb{P}[Z = z_j | Z^* = z^*]. \end{aligned}$$

Therefore, $\mathcal{R}(S^*)$ is generated by the functions

$$r_j(z^*) = \Pi_{\overline{T(\lambda_0)}} \mathbb{P}[Z = z_j | Z^* = z^*], \text{ for } j = 1, \dots, m.$$

Therefore, $\mathcal{R}(S^*)$ is finite-dimensional, and then closed. ■

Proof of Proposition 4.4: Define

$$\begin{aligned} b(s) &= \mathbb{E}[g(Y_i = 1, X_i) | \beta_i = s] \\ &= \int 1(x's \geq 0) g(1, x) dv_X(x). \end{aligned}$$

I prove that b is continuous and by compactness of the sphere is therefore uniformly continuous. Since the halfspaces $1(x's \geq 0)$ and $1(x's_0 \geq 0)$ intersect in sets having surface measure of order $|s - s_0|$, it follows from the absolute continuity of the angular component of X that

$$|b(s) - b(s_0)| = O(|s - s_0|).$$

When $x = (1, \tilde{x})$, then

$$b(s) = \int 1(\tilde{x}'s \geq -s_1) g(1, 1, \tilde{x}) dv_X(\tilde{x}),$$

and the absolute continuity in s_1 follows from the integrability of $g(1, 1, \tilde{x})$. ■

Proof of Proposition 5.1: For the functional $\phi(\lambda) = \theta$ it holds $\mathcal{N}(\dot{\phi}) = \{(b_\theta, b_\eta) : b_\theta = 0\}$. Then, by orthogonality

$$\begin{aligned} \mathcal{N}(S) &= \left\{ (b_\theta, b_\eta) : \int (\dot{l}_\theta b_\theta + \dot{l}_\eta b_\eta)^2 d\mathbb{P}_{\theta_0, \eta_0} = 0 \right\} \\ &= \left\{ (b_\theta, b_\eta) : \int (\tilde{l}_\theta b_\theta + \Pi_{\overline{\mathcal{R}(\dot{l}_\eta)}} \dot{l}_\theta b_\theta + \dot{l}_\eta b_\eta)^2 d\mathbb{P}_{\theta_0, \eta_0} = 0 \right\} \\ &= \left\{ (b_\theta, b_\eta) : \int (\tilde{l}_\theta b_\theta)^2 d\mathbb{P}_{\theta_0, \eta_0} = 0, \int (\Pi_{\overline{\mathcal{R}(\dot{l}_\eta)}} \dot{l}_\theta b_\theta + \dot{l}_\eta b_\eta)^2 d\mathbb{P}_{\theta_0, \eta_0} = 0 \right\} \\ &= \left\{ (b_\theta, b_\eta) : b'_\theta \tilde{I}_\theta b_\theta = 0, \Pi_{\overline{\mathcal{R}(\dot{l}_\eta)}} \dot{l}_\theta b_\theta = -\dot{l}_\eta b_\eta \right\}. \end{aligned}$$

Then, if $\tilde{I}_\theta > 0$, then we have $\mathcal{N}(S) \subset \mathcal{N}(\dot{\phi})$. If $\tilde{I}_\theta = 0$, then there are two cases: (1) $\dot{l}_\theta \in \mathcal{R}(\dot{l}_\eta)$ and (2) $\dot{l}_\theta \in \overline{\mathcal{R}(\dot{l}_\eta)} \setminus \mathcal{R}(\dot{l}_\eta)$. In case (1) the identification condition $\mathcal{N}(S) \subset \mathcal{N}(\dot{\phi})$ does not hold, as we can find $b_\theta \neq 0$ such that $\dot{l}_\theta b_\theta = -\dot{l}_\eta b_\eta$ (so $(b_\theta, b_\eta) \in \mathcal{N}(S)$ but $(b_\theta, b_\eta) \notin \mathcal{N}(\dot{\phi})$). In case (2) $\mathcal{N}(S) \subset \mathcal{N}(\dot{\phi})$ holds even though there is zero information for the parameter. Thus, if \dot{l}_θ is at the boundary of $\mathcal{R}(\dot{l}_\eta)$ we have “irregular identification”. ■

Proof of Proposition 5.2: Note that for the functional $\phi(\lambda) = \chi(\eta)$, where $\chi : H \mapsto \mathbb{R}$ is a linear continuous functional with

$$\chi(b_\eta) = \langle b_\eta, r_\chi \rangle_H,$$

it holds that $\mathcal{N}(\dot{\phi}) = \{(b_\theta, b_\eta) : \langle b_\eta, r_\chi \rangle_H = 0\}$. Therefore, $\mathcal{N}(S) \subset \mathcal{N}(\dot{\phi})$ if $\tilde{I}_\theta b_\theta = 0$ and $\Pi_{\overline{\mathcal{R}(\dot{l}_\eta)}} \dot{l}_\theta b_\theta = -\dot{l}_\eta b_\eta$ implies $\langle b_\eta, r_\chi \rangle_H = 0$. If $\tilde{I}_\theta > 0$, then $(b_\theta, b_\eta) \in \mathcal{N}(S)$ iff $b_\theta = 0$ and $0 = \dot{l}_\eta b_\eta$. Therefore,

$(b_\theta, b_\eta) \in \mathcal{N}(\dot{\phi})$ if $\mathcal{N}(\dot{l}_\eta) \subset \mathcal{N}(\chi)$, which is equivalent to $r_\chi \in \overline{\mathcal{R}(\dot{l}_\eta^*)}$. If $\tilde{I}_\theta = 0$, there are two cases (i) $\dot{l}_\theta \in \mathcal{R}(\dot{l}_\eta)$ and (ii) $\dot{l}_\theta \in \overline{\mathcal{R}(\dot{l}_\eta)} \setminus \mathcal{R}(\dot{l}_\eta)$. In case (i), $\dot{l}_\theta b_\theta = -\dot{l}_\eta b_\eta$ and we can take $b_\theta = -1$, and for all such b_η it must hold that $\langle b_\eta, r_\chi \rangle_H = 0$. This is equivalent to $\langle b_\eta^0, r_\chi \rangle_H = 0$ and $r_\chi \in \overline{\mathcal{R}(\dot{l}_\eta^*)}$. In case (ii) $0 = \dot{l}_\eta b_\eta$ must imply that $(b_\theta, b_\eta) \in \mathcal{N}(\dot{\phi})$, which holds if $\mathcal{N}(\dot{l}_\eta) \subset \mathcal{N}(\chi)$ or equivalently $r_\chi \in \overline{\mathcal{R}(\dot{l}_\eta^*)}$. Therefore, if $\dot{l}_\theta \notin \mathcal{R}(\dot{l}_\eta)$ ($\tilde{I}_\theta > 0$ or case (ii) above) then $\mathcal{N}(S) \subset \mathcal{N}(\dot{\phi})$ holds iff $r_\chi \in \overline{\mathcal{R}(\dot{l}_\eta^*)}$; (ii) if $\dot{l}_\theta \in \mathcal{R}(\dot{l}_\eta)$ (case (i) above) then $\mathcal{N}(S) \subset \mathcal{N}(\dot{\phi})$ holds iff $r_\chi \in \overline{\mathcal{R}(\dot{l}_\eta^*)}$ and $\langle b_\eta^0, r_\chi \rangle_H = 0$. ■

Proof of Theorem 5.3: For $\varepsilon < I_{\theta, \rho}^{1/2}$, we can choose $\delta > 0$ such that for all $\lambda \in \mathcal{B}_\delta(\lambda_0)$ with $\theta \neq \theta_0$

$$\begin{aligned} \frac{\|(f_\lambda - f_{\lambda_0})/f_{\lambda_0} - S(\lambda - \lambda_0)\|}{\|S(\lambda - \lambda_0)\|} &= \frac{\|(f_\lambda - f_{\lambda_0})/f_{\lambda_0} - S(\lambda - \lambda_0)\|}{|\theta - \theta_0|^\rho} \frac{|\theta - \theta_0|^\rho}{\|S(\lambda - \lambda_0)\|} \\ &\leq \varepsilon \times I_{\theta, \rho}^{-1/2} \\ &< 1, \end{aligned} \tag{20}$$

where we have used Assumption 2 and the definition of the generalized Fisher information. The inequality (20) implies that $\|(f_\lambda - f_{\lambda_0})/f_{\lambda_0}\| \neq 0$, or equivalently $f_\lambda \neq f_{\lambda_0}$. That is, local identification holds. ■

Proof of Corollary 5.1: Follows from Theorem 5.3 and $\tilde{I}_\theta \equiv I_{\theta, 1} \leq I_{\theta, \rho}$ for $1 < \rho < \infty$. ■

Proof of Theorem 6.1: The proof follows closely that of Chamberlain (2010, Theorem 2). I first show the model is mean-square differentiable along paths of the form

$$\eta_t = \eta_0(1 + tb_\eta),$$

where $\eta_0 \in H$, $t \in (0, \varepsilon)$, $\varepsilon > 0$, and $b_\eta : \mathbb{R}^K \rightarrow \mathbb{R}$ is a bounded, measurable function such that

$$\int \eta_0(\beta) b_\eta(\beta) d\beta = 0.$$

Define the parametric likelihood $f_\gamma(y, x) = \int f_{y/x, \beta}(y; \theta) \eta_t(\beta) d\beta$, with $\gamma = (\theta, t)$, and apply the mean-value theorem at $\gamma_0 = (\theta_0, 0)$ to obtain

$$f_\gamma^{1/2}(y, x) - f_{\gamma_0}^{1/2}(y, x) = \frac{\partial f_{\gamma_0}^{1/2}(y, x)}{\partial \gamma'} (\gamma - \gamma_0) + r(y, x; \gamma),$$

where

$$r(y, x; \gamma) = \left[\frac{\partial f_{\tilde{\gamma}}^{1/2}(y, x)}{\partial \gamma'} - \frac{\partial f_{\gamma_0}^{1/2}(y, x)}{\partial \gamma'} \right] (\gamma - \gamma_0)$$

and $\tilde{\gamma}$ is on the line segment joining γ and γ_0 . Note that for the mixed Logit it holds

$$\frac{\partial f_{y/x, \beta}(y; \theta_0)}{\partial \theta_{0y}} = f_{y/x, \beta}(y; \theta_0) (1 - f_{y/x, \beta}(y; \theta_0)).$$

Thus, the partial derivatives $\partial f_\gamma^{1/2}(y, x)/\partial \gamma$ are well-defined and are continuous in γ . Then,

$$\frac{r(y, x; \gamma)}{|\gamma - \gamma_0|^2} \leq \left| \frac{\partial f_{\tilde{\gamma}}^{1/2}(y, x)}{\partial \gamma'} - \frac{\partial f_{\gamma_0}^{1/2}(y, x)}{\partial \gamma'} \right|^2 \rightarrow 0$$

as $\gamma \rightarrow \gamma_0$ (μ -a.e) and by dominated convergence

$$\int \frac{r^2(y, x; \gamma)}{|\gamma - \gamma_0|^2} \mu(dz) \rightarrow 0.$$

This verifies the mean-square differentiability of the model.

Thus, the score operator is given by

$$S(b_\theta, b_\eta) = i'_\theta b_\theta + i_\eta b_\eta,$$

where

$$i_\theta = \frac{1}{f_{\lambda_0}(y, x)} \int \frac{\partial f_{y/x, \beta}(y; \theta_0)}{\partial \theta} \eta_0(\beta) d\beta$$

and

$$\begin{aligned} i_\eta b_\eta(y, x) &= \frac{1}{f_{\lambda_0}(y, x)} \int f_{y/x, \beta}(y; \theta_0) b_\eta(\beta) \eta_0(\beta) d\beta \\ &= \mathbb{E}[b_\eta(\beta) | Y = y, X = x], \end{aligned}$$

where $b_\theta \in \mathbb{R}^J$, $b_\eta \in L_2^0(G_0)$ and G_0 is the distribution of β .

Next, I show that $\mathcal{R}(i_\eta) = L_2^0$ and as a result θ_{0y} has zero information for $y = 1, \dots, J$. The condition $\mathcal{R}(i_\eta) = L_2^0$ means that for any $g \in L_2^0$ and $\varepsilon > 0$, we can find $b_\eta \in L_2^0(G_0)$ such that

$$\|g - i_\eta b_\eta\| < \varepsilon.$$

Since $g \in L_2^0$, this will hold if we can choose for any $\varepsilon_1 > 0$, a compact set $B \subset \mathcal{X}$ such that $v_X(B) > 1 - \varepsilon_1$ and

$$\sum_{j=0}^J \int_B \left(g(j, x) - \frac{1}{f_{\lambda_0}(j, x)} \int f_{y/x, \beta}(j; \theta_0) b_\eta(\beta) \eta_0(\beta) d\beta \right)^2 v_X(dx) < \varepsilon_1. \quad (21)$$

Let $r(x)$ denote the $(J+1) \times 1$ vector with elements $g(j, x) f_{\lambda_0}(j, x)$ for $j = 0, \dots, J$. Note that by the zero mean property, $l'r(x) = 0$, where l is the $(J+1) \times 1$ vector of ones. Let $a(x, \beta)$ denote the $(J+1) \times 1$ vector with elements $f_{y/x, \beta}(j; \theta_0)$ for $j = 0, \dots, J$. Then, since $f_{\lambda_0}(j, x)$ is bounded away from zero on $x \in B$, (21) is satisfied if for all $\varepsilon_2 > 0$ there is a $b_\eta(\beta)$ with the properties above such that

$$\left| r(x) - \int a(x, \beta) b_\eta(\beta) \eta_0(\beta) d\beta \right| < \varepsilon_2. \quad (22)$$

Suppose that for all x ν_X -a.s, there exists points $\beta_j(x) \in \mathbb{R}^K$ with

$$H(x) = [a(x, \beta_0(x)), \dots, a(x, \beta_J(x))]$$

non-singular. Then, for each such x , there is a neighborhood C_x of x such that

$$[a(\tilde{x}, \beta_0(x)), \dots, a(\tilde{x}, \beta_J(x))]$$

is non-singular for all \tilde{x} in the closure of C_x . The C_x provide an open cover of a compact set B with $\nu_X(B) > 1 - \varepsilon_1$. Hence, there is a finite subcover, and we can partition B into Borel subsets D_1, \dots, D_m and choose the $\beta_j(x)$ to be simple functions and such that $H(x)$ has a determinant bounded away from zero for all $x \in B$. Then, we can find $c_j(x)$ such that for all $x \in B$

$$r(x) = \sum_{j=0}^J a(\tilde{x}, \beta_j(x)) c_j(x).$$

Then, (22) holds by setting for all $x \in B$

$$b_\eta(\beta) = \frac{1}{\eta_0(\beta)} \sum_{j=0}^J 1(|\beta - \beta_j(x)| < \delta) c_j(x) / 2\delta,$$

for δ sufficiently small.

We conclude that $\overline{\mathcal{R}(\dot{l}_\eta)} = L_2^0$ unless, for all x in a set with positive ν_X measure, $\{a(x, \beta) : \beta \in \mathbb{R}^K\}$ lies in a proper linear subspace of \mathbb{R}^{J+1} . Then, for each such x , there exists a nonzero vector with components φ_j such that

$$\sum_{j=0}^J \varphi_j f_{y/x, \beta}(j; \theta_0) = 0,$$

or

$$\varphi_0 + \sum_{j=1}^J \varphi_j \exp(\theta_{0j} + x'_j \beta) = 0,$$

which is not possible by Assumption 3 since by differentiating with respect to β the last expression one obtains $\sum_{j=0}^J \varphi_j x_j \exp(\theta_{0j} + x'_j \beta) = 0$, which evaluated at $x_{-j} = 0$ simplifies to $\varphi_j x_j \exp(\theta_{0j} + x'_j \beta) = 0$. Then, use that $\mathbb{E}[x_j x'_j]$ is positive definite to conclude $\varphi_j = 0$, for all $j = 0, \dots, J$. ■

8.2 Relation with Singular Value Decomposition

In linear models, the key rank condition $r_\phi \in \overline{\mathcal{R}(S^*)}$ was necessary and sufficient for identification. I introduce some tools that allow to interpret conditions such as $r_\phi \in \overline{\mathcal{R}(S^*)}$ or the stronger $r_\phi \in \mathcal{R}(S^*)$. These tools, however, require an additional assumption, which often holds in most applications.

Assumption C: The score operator S is compact.

Assumption C guarantees the existence of a sequence $\{\lambda_j, \varphi_j, \psi_j\}_{j=1}^\infty$ such that (cf. Kress, 1999, Theorem 15.16)

$$S\varphi_j = \lambda_j \psi_j \quad \text{and} \quad S^* \psi_j = \lambda_j \varphi_j. \quad (23)$$

This is the so called singular value decomposition of S . The elements $\{\varphi_j\}_{j=1}^\infty$ and $\{\psi_j\}_{j=1}^\infty$ are complete orthonormal bases for $\overline{\mathcal{R}(S^*)}$ and $\overline{\mathcal{R}(S)}$, respectively, and the singular values λ_j are the squared-root eigenvalues of the information operator $I_{\lambda_0} := S^* S : T(\lambda_0) \mapsto T(\lambda_0)$. Furthermore, defining for $\beta \in \mathbb{R}$,

$$\mathcal{M}_\beta := \left\{ b \in T(\lambda_0) \text{ such that } \|b\|_\beta^2 := \sum_{j=1}^\infty \lambda_j^{-2\beta} \langle b, \varphi_j \rangle_{\mathbf{H}}^2 < \infty \right\},$$

it is well known (see e.g. Carrasco, Florens and Renault 2006) that

$$\overline{\mathcal{R}(S^*)} \equiv \mathcal{M}_0 = \left\{ b \in T(\lambda_0) \text{ such that } \sum_{j=1}^{\infty} \langle b, \varphi_j \rangle_{\mathbf{H}}^2 < \infty \right\},$$

whereas

$$\mathcal{R}(S^*) \equiv \mathcal{M}_1 = \left\{ b \in T(\lambda_0) \text{ such that } \sum_{j=1}^{\infty} \lambda_j^{-2} \langle b, \varphi_j \rangle_{\mathbf{H}}^2 < \infty \right\}.$$

With this notation, functionals with $r_\phi \in \overline{\mathcal{R}(S^*)} \setminus \mathcal{R}(S^*)$ correspond to those such that $\|r_\phi\|_\beta < \infty$ for $0 \leq \beta < 1$, whereas positive information corresponds to $\|r_\phi\|_1 < \infty$ (see below for a proof).

Any element $b \in T(\lambda_0)$ has the singular value expansion (cf. Kress, 1999, Theorem 15.16)

$$b = \sum_{j=1}^{\infty} \langle b, \varphi_j \rangle_{\mathbf{H}} \varphi_j + \Pi_{\mathcal{N}(S)} b,$$

which implies under identification

$$\dot{\phi}(b) = \sum_{j=1}^{\infty} \langle b, \varphi_j \rangle_{\mathbf{H}} \langle r_\phi, \varphi_j \rangle_{\mathbf{H}}$$

and

$$Sb = \sum_{j=1}^{\infty} \lambda_j \langle b, \varphi_j \rangle_{\mathbf{H}} \psi_j.$$

By Cauchy-Schwarz, for $b \in T(\lambda_0)$,

$$\begin{aligned} |\dot{\phi}(b)| &\leq \left(\sum_{j=1}^{\infty} \lambda_j^{-2} \langle r_\phi, \varphi_j \rangle_{\mathbf{H}}^2 \right)^{1/2} \left(\sum_{j=1}^{\infty} \lambda_j^2 \langle b, \varphi_j \rangle_{\mathbf{H}}^2 \right)^{1/2} \\ &= \|r_\phi\|_1 \|Sb\|. \end{aligned}$$

Therefore, regular identification can be interpreted as continuity of $|\dot{\phi}(b)|$ with respect to the semi-norm $\|Sb\|$. This also shows one of the implications of Theorem 4.1 in van der Vaart (1991)

$$I_\phi = \inf_{b \in T(\lambda_0)} \frac{\|Sb\|^2}{[\dot{\phi}(b)]^2} \geq \frac{1}{\|r_\phi\|_1} > 0.$$

More generally, by Holder inequality, for any $0 < \beta \leq 1$ and for all b with $\|b\|_{\mathbf{H}} \leq 1$,

$$\begin{aligned} |\dot{\phi}(b)| &\leq \left(\sum_{j=1}^{\infty} \lambda_j^{-2\beta} \langle r_\phi, \varphi_j \rangle_{\mathbf{H}}^2 \right)^{1/2} \left(\sum_{j=1}^{\infty} \lambda_j^{2\beta} \langle b, \varphi_j \rangle_{\mathbf{H}}^2 \right)^{1/2} \\ &\leq \|r_\phi\|_\beta \left(\sum_{j=1}^{\infty} \lambda_j^2 \langle b, \varphi_j \rangle_{\mathbf{H}}^2 \right)^{\beta/2} \left(\sum_{j=1}^{\infty} \langle b, \varphi_j \rangle_{\mathbf{H}}^2 \right)^{(1-\beta)/2} \\ &\leq \|r_\phi\|_\beta \|Sb\|^\beta. \end{aligned}$$

Thus, we have the following result.

Theorem 8.1 *Let Assumption 1 and Assumption C hold. Then (i) $\phi(\lambda_0)$ is regularly identified iff $\|r_\phi\|_1 < \infty$; (ii) $\phi(\lambda_0)$ is irregularly identified if $\|r_\phi\|_1 = \infty$ but $\|r_\phi\|_\beta < \infty$ for $0 < \beta < 1$.*

It is known that in many cases, bounds on the Fourier coefficients $\langle b, \varphi_j \rangle_{\mathbf{H}}^2$ correspond to imposing smoothing conditions on b (see Kress 1999, Chapter 8). Hence, in these cases one can index the level of irregularity by the smoothness of the influence function r_ϕ .

8.3 Linear Models in Nuisance Parameters without Completeness

Define the nuisance score operator

$$\dot{l}_{\eta(\theta)} b_\eta = \frac{f_{\theta, \eta_0 + b_\eta} - f_{\theta, \eta_0}}{f_{\theta_0, \eta_0}}, \quad (24)$$

and the (negative) approximated score for θ as

$$s_\theta = \frac{f_{\theta_0, \eta_0} - f_{\theta, \eta_0}}{f_{\theta_0, \eta_0}}.$$

I drop the dependence on θ_0 and denote $\dot{l}_\eta \equiv \dot{l}_{\eta(\theta_0)}$. Define the (negative) approximated efficient score $\tilde{s}_\theta := s_\theta - \Pi_{\mathcal{R}(\dot{l}_\eta)} s_\theta$, and the approximated square-root Fisher Information

$$G(\theta) = \|\tilde{s}_\theta\|.$$

Let Ψ be the class of measurable functions $\psi : [0, \infty) \rightarrow [0, \infty)$ that are non-decreasing, right continuous at 0 and with $\psi(0) = 0$. Then, consider the following assumption.

Assumption D: (i) The map $\dot{l}_{\eta(\theta)} : T(\eta_0) \subseteq \mathcal{H} \mapsto L_2$ is linear for each θ in a neighborhood of θ_0 (ii) there exists a positive constant C such that $G(\theta) > C\psi(|\theta - \theta_0|)$ in a neighborhood of θ_0 , where $\psi \in \Psi$.

Assumption D(i) holds for many models of interest. Assumption D(ii) follows from conditions on the derivative of $G(\theta)$ at θ_0 . For example, if $G(\theta)$ is differentiable at θ_0 with full rank derivative at θ_0 , then Assumption D(ii) holds with $\psi(\epsilon) = \epsilon$. This corresponds to the case of regular local identification. A necessary condition for Assumption D(ii) is that $\mathcal{N}(\dot{l}_{\eta(\theta)}^*) \neq 0$, since otherwise $G(\theta) = 0$.

Theorem 8.2 *Let Assumption D hold. Then, θ is locally identified at θ_0 .*

Proof of Theorem 8.2: Write

$$\begin{aligned} \frac{f_{\theta, \eta} - f_{\theta_0, \eta_0}}{f_{\theta_0, \eta_0}} &= \frac{f_{\theta, \eta} - f_{\theta, \eta_0}}{f_{\theta_0, \eta_0}} - \frac{f_{\theta_0, \eta_0} - f_{\theta, \eta_0}}{f_{\theta_0, \eta_0}} \\ &= \dot{l}_{\eta(\theta)} b_\eta - s_\theta. \end{aligned}$$

Note that by standard least squares theory for all $b_\eta \in T(\eta_0)$, and all θ in a neighborhood of θ_0 ,

$$\begin{aligned} \|\dot{l}_{\eta(\theta)} b_\eta - s_\theta\| &\geq \|\Pi_{\mathcal{R}(\dot{l}_{\eta(\theta)})} s_\theta - s_\theta\| \\ &> C\psi(|\theta - \theta_0|). \end{aligned}$$

This inequality implies local identification. ■

8.4 General Nonlinear Models

The following modulus of continuity is shown to be useful for the study of identification

$$\varpi(\epsilon) = \sup_{\lambda \in \mathcal{B}_\delta(\lambda_0) : \|(f_\lambda - f_{\lambda_0})f_{\lambda_0}^{-1}\| \leq \epsilon} |\phi(\lambda) - \phi(\lambda_0)|. \quad (25)$$

I drop the dependence of $\varpi(\epsilon)$ on δ for simplicity of notation. Lemma 8.4 below shows that $\varpi(\epsilon) \downarrow 0$ as $\epsilon \downarrow 0$ is sufficient for local identification of $\phi(\lambda_0)$. A related modulus of continuity was introduced in Donoho and Liu (1987) for the purpose of obtaining bounds on the optimal rate of convergence for functionals of a density (they assume identification and use the Hellinger metric). Using $\|(f_\lambda - f_{\lambda_0})f_{\lambda_0}^{-1}\|$ is convenient because we can exploit simultaneously the linearity of certain models and the Hilbert space structure.

Lemma *If there exists $\delta > 0$ such that $\varpi(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$ then $\phi(\lambda_0)$ is locally identified.*

Proof of Lemma 8.4: Suppose that $\phi(\lambda_0)$ is not locally identified. Then, for all $\delta > 0$, we can find a $\lambda^* \in \Lambda_\delta(\lambda_0)$ such that $\|(f_{\lambda^*} - f_{\lambda_0})/f_{\lambda_0}\| = 0$, and therefore, for all $\epsilon > 0$,

$$\varpi(\epsilon) \geq |\phi(\lambda^*) - \phi(\lambda_0)| > 0,$$

showing that $\varpi(\epsilon)$ does not converge to zero as $\epsilon \rightarrow 0$. ■

The following result provides a general local identification result under the conditions in

Assumption N: *For all $\varepsilon > 0$, there exists $\delta > 0$, $\psi_1, \psi_2 \in \Psi$, and a continuous linear operator $S : T(\lambda_0) \subseteq H \mapsto L_2$, such that for all $\lambda = (\theta, \eta) \in \mathcal{B}_\delta(\lambda_0)$,*

(i)

$$\|(f_\lambda - f_{\lambda_0})/f_{\lambda_0} - S(\lambda - \lambda_0)\| < \varepsilon \psi_1(\|\lambda - \lambda_0\|_{\mathbf{H}});$$

(ii)

$$|\phi(\lambda) - \phi(\lambda_0)| \leq \psi_2(\|\lambda - \lambda_0\|_{\mathbf{H}});$$

(iii) and

$$\inf_{\lambda \in \mathcal{B}_\delta(\lambda_0)} \frac{\|S(\lambda - \lambda_0)\|}{\psi_1(\|\lambda - \lambda_0\|_{\mathbf{H}})} > 0.$$

Assumption N(i) and N(ii) are mild smoothness conditions that often hold in applications. Condition N(iii) is a positive nonparametric generalized information condition. Then, I have the following

Theorem 8.3 *Let Assumption N hold. Then, $\phi(\lambda)$ is locally identified at $\phi(\lambda_0)$.*

Proof of Theorem 8.3: Assumptions N(i-ii) imply that if $\|(f_\lambda - f_{\lambda_0})f_{\lambda_0}^{-1}\| \leq \epsilon$ then we can find a positive constant C and $0 < \varepsilon < C$ such that for all $\lambda = (\theta, \eta) \in \mathcal{B}_\delta(\lambda_0)$,

$$C\psi_1(\|\lambda - \lambda_0\|_{\mathbf{H}}) \leq \|S(\lambda - \lambda_0)\| \leq \varepsilon\psi_1(\|\lambda - \lambda_0\|_{\mathbf{H}}) + \epsilon,$$

which in turn implies

$$\psi_1(\|\lambda - \lambda_0\|_{\mathbf{H}}) \leq \frac{\epsilon}{C - \varepsilon}.$$

Hence, by Assumption N(ii)

$$\begin{aligned} \varpi(\epsilon) &= \sup_{\lambda \in \mathcal{B}_\delta(\lambda_0): \|(f_\lambda - f_{\lambda_0})f_{\lambda_0}^{-1}\| \leq \epsilon} |\phi(\lambda) - \phi(\lambda_0)|, \\ &\leq \sup_{\lambda \in \mathcal{B}_\delta(\lambda_0): \psi_1(\|\lambda - \lambda_0\|_{\mathbf{H}}) \leq \frac{\epsilon}{C - \varepsilon}} \psi_2(\|\lambda - \lambda_0\|_{\mathbf{H}}) \\ &\leq \psi_2\left(\psi_1^{-1}\left(\frac{\epsilon}{C - \varepsilon}\right)\right) \\ &\rightarrow 0 \text{ as } \epsilon \rightarrow 0. \end{aligned}$$

Thus, the Theorem follows from Lemma 8.4. ■

Assumption N implies that λ_0 is locally identified, which might be strong for some applications. Relaxing this condition in a general nonlinear settings turns out to be a rather delicate problem. The main issue is that $\|(f_\lambda - f_{\lambda_0})/f_{\lambda_0} - S(\lambda - \lambda_0)\|$ is generally not continuous with respect to $|\phi(\lambda) - \phi(\lambda_0)|$. To overcome this problem, I consider a profiling approach. For any $\lambda \in \Lambda$, write $\lambda = \lambda_0 + \lambda_r + \lambda_{r\perp}$, where $\lambda_r = \langle \lambda - \lambda_0, r_\phi \rangle_{\mathbf{H}} \langle r_\phi, r_\phi \rangle_{\mathbf{H}}^{-1} r_\phi \equiv tr_\phi$ and $\lambda_{r\perp} = \lambda - \lambda_0 - \lambda_r$. Define

$$\lambda_{r\perp}^*(\lambda_r) = \arg \min_{\lambda_{r\perp}: \|\lambda_{r\perp}\|_{\mathbf{H}} \leq \delta} \left\| (f_{\lambda_0 + \lambda_r + \lambda_{r\perp}} - f_{\lambda_0}) / f_{\lambda_0} \right\|,$$

which is the least favorable direction for a fixed λ_r . Let $m(\lambda_r) = (f_{\lambda_0 + \lambda_r + \lambda_{r\perp}^*(\lambda_r)} - f_{\lambda_0}) / f_{\lambda_0}$.

Assumption P: Suppose $\varphi(t) \equiv m(tr_\phi)$ is differentiable at $t = 0$ with derivative $\dot{\varphi}$. If for all $\varepsilon > 0$ there exists $\delta > 0$, $\rho \geq 1$, such that for all $|t| \leq \delta$: $\|\varphi(t) - \varphi(0) - \dot{\varphi}t\| \leq \varepsilon|t|^\rho$ and for a positive constant C , $\|\dot{\varphi}t\| \geq C|t|^\rho$. In addition, Assumptions 1(i) and 1(iv) hold.

Theorem 8.4 Under Assumption P, $\phi(\lambda_0)$ is locally identified.

Proof of Theorem 8.4: For a sufficiently small $\delta > 0$

$$\begin{aligned} \|(f_{\theta, \eta} - f_{\theta_0, \eta_0}) / f_{\theta_0, \eta_0}\| &\geq \|m(\lambda_r) - m(0)\| \\ &\geq \|\dot{\varphi}t\| - \varepsilon|\lambda_r|^\rho \\ &> C|\lambda_r|^\rho \\ &\equiv C_r |\phi(\lambda) - \phi(\lambda_0)|^\rho \end{aligned}$$

for all $\lambda \in \mathcal{B}_\delta(\lambda_0)$ and a positive constant C_r . That is, local identification of $\phi(\lambda_0)$ holds. ■

One setting where Assumption P is easy to check is that of models that are nonlinear in the parameter of interest but linear in nuisance parameters, as shown in the proof of Theorem 8.2 above.

References

- AN, Y. AND Y. HU (2012), “Well-posedness of measurement error models for self-reported data,” *Journal of Econometrics*, 168, 259-269.
- ALVAREZ, F., BOROVÍČKOVÁ, K. AND R. SHIMER (2016): “Decomposing Duration Dependence in a Stopping Time Model,” manuscript.
- ANDREWS, D.W.K (2011): “Examples of L2-Complete and Boundedly-Complete Distributions,” Cowles Foundation for Research in Economics.
- ANDREWS, D.W.K AND M. A. SCHAFGANS (1998): “Semiparametric Estimation of the Intercept of a Sample Selection Model,” *Review of Economic Studies*, 65, 497-517.
- ARELLANO, M. AND S. BONHOMME (2011): “Nonlinear Panel Data Analysis,” *Annual Review of Economics*, 3, 395-424.
- BAJARI, P., J. HAHN, H. HONG, AND G. RIDDER (2011): “A Note on Semiparametric Estimation of Finite Mixtures of Discrete Choice Models With Application to Game Theoretic Models,” *International Economic Review*, 52, 807-824.
- BEKKER, P. AND WANSBEEK, T. (2001): “Identification in parametric models,” in B. Baltagi, ed., ‘*Companion to Theoretical Econometrics*’, Blackwell Companions to Contemporary Economics, Basil Blackwell, Oxford, U.K., chapter 7, pp. 144-161.
- BEGUN, J. M., W. J. HALL, W. M. HUANG, AND J. A. WELLNER (1983): “Information and Asymptotic Efficiency in Parametric-Nonparametric Models,” *Annals of Statistics*, 11, 432-452.
- BICKEL, P. J., C. A. J. KLASSEN, Y. RITOV, AND J. A. WELLNER (1998): *Efficient and Adaptive Estimation for Semiparametric Models*. New York: Springer-Verlag.
- BLUNDELL, R., X. CHEN, AND D. KRISTENSEN (2007): “Semi-nonparametric IV Estimation of Shape-invariant Engel Curves,” *Econometrica*, 75, 1613–1670.
- BONHOMME, S. (2011): “Panel Data, Inverse Problems, and the Estimation of Policy Parameters,” Unpublished manuscript.
- BONHOMME, S. (2012): “Functional Differencing,” *Econometrica*, 80, 1337-1385.
- BOYD, J.H., AND MELLMAN, R.E. (1980): “Effect of fuel economy standards on the US automotive market: an hedonic demand analysis,” *Transportation Research B* 14(5), 367-378.
- BUCHINSKY, M., J. HAHN, AND K. KIM (2010): “Semiparametric Information Bound of Dynamic Discrete Choice Models,” *Economics Letters*, 108, 109-112.
- CARDELL, N.S. AND DUNBAR, F.C. (1980): “Measuring the societal impacts of automobile downsizing,” *Transportation Research B* 14(5), 423-434.

- CARRASCO, M., J. P. FLORENS, AND E. RENAULT (2006): “Linear Inverse Problem in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization,” in *Handbook of Econometrics*, vol. 6, ed. by J. J. Heckman and E. E. Leamer. Amsterdam: North-Holland, 5633–5751.
- CATTANEO, M. AND ESCANCIANO, J.C. (2017): “Regression Discontinuity Designs: Theory and Applications,” Volume 38 of *Advances in Econometrics*, in print.
- CHAMBERLAIN, G. (1986): “Asymptotic Efficiency in Semi-Parametric Models with Censoring,” *Journal of Econometrics*, 34, 305-334.
- CHAMBERLAIN, G. (1992): “Efficiency Bounds for Semiparametric Regression,” *Econometrica*, 60, 567-596.
- CHAMBERLAIN, G. (2010): “Binary Response Models for Panel Data: Identification and Information,” *Econometrica*, 78, 159-168.
- CHEN, X., V. CHERNOZHUKOV, S. LEE, AND W. NEWEY (2014), “Identification in Semiparametric and Nonparametric Conditional Moment Models,” *Econometrica*, 82, 785-809.
- CHEN, X. AND S. C. LUDVIGSON (2009): “Land of addicts? An Empirical Investigation of Habit-Based Asset Pricing Models,” *Journal of Applied Econometrics*, 24, 1057-1093.
- CHEN, X., POUZO, D. (2015) “Sieve quasi likelihood ratio inference on semi/nonparametric conditional moment models,” *Econometrica* 83, 1013–1079.
- CHEN, X., AND A. SANTOS (2015): “Overidentification in Regular Models,” working paper.
- CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I., HAHN, J., AND NEWEY, W.K (2013): “Average and Quantile Effects in Nonseparable Panel Models,” *Econometrica*, 81, 535-580.
- DAVEZIES, L. (2015): “On the Existence of \sqrt{N} -consistent Estimators of Linear Functionals in Non-Parametric IV Models”, unpublished manuscript.
- D’HAULTFOEUILLE, X. (2011): “On the Completeness Condition in Nonparametric Instrumental Problems,” *Econometric Theory*, 1, 1-12.
- DONOHU, D.L. AND LIU, R.C. (1987): “Geometrizing rates of convergence, I.” Technical Report 137, Dept. Statistics, Univ. California, Berkeley.
- DUFOR, J.M. AND LIANG, X. (2014): “Necessary and sufficient conditions for nonlinear parametric function identification”, unpublished manuscript.
- ELBERS, C. AND G. RIDDER (1982): “True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model,” *Review of Economic Studies*, 49, 403-409.

- ESCANCIANO, J.C., S. HODERLEIN, A. LEWBEL, O. LINTON AND S. SRISUMA, (2015): "Nonparametric Euler Equation Identification and Estimation," unpublished manuscript.
- FISHER, F. M. (1966), *The Identification Problem in Econometrics*, McGraw-Hill, New York.
- FOX, J. T., K.-I. KIM, S. P. RYAN, AND P. BAJARI (2012): "The Random Coefficients Logit Model Is Identified," *Journal of Econometrics*, 166 (2), 204-212.
- GAUTIER, E. AND Y. KITAMURA (2013): "Nonparametric estimation in random coefficients binary choice models," *Econometrica*, 81, 581-607.
- GRAHAM, B. W., AND J. L. POWELL (2012): "Identification and Estimation of Average Partial Effects in 'Irregular' Correlated Random Coefficient Panel Data Models," *Econometrica*, 80, 2105-2152.
- HAHN, J. (1994): "The Efficiency Bound of the Mixed Proportional Hazard Model," *Review of Economic Studies*, 61, 607-629.
- HAHN, J. (2001): "The Information Bound of a Dynamic Panel Logit Model with Fixed Effects," *Econometric Theory*, 5, 913-932.
- HECKMAN, J. J. (1990), "Varieties of selection bias", *The American Economic Review* 80, 313-318.
- HECKMAN, J.J. AND B. SINGER (1984a): "The Identifiability of the Proportional Hazard Model", *Review of Economic Studies*, 51,231-241.
- HECKMAN, J.J. AND B. SINGER (1984b): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52, 271-320.
- HODERLEIN, S., L. NESHEIM, AND A. SIMONI (2012): "Semiparametric estimation of random coefficients in structural economic models," Working paper.
- HODERLEIN S. AND SHERMAN, B. (2015): "Identification and estimation in a correlated random coefficients binary response model," *Journal of Econometrics*, 188, 135-149.
- HONORÉ, B., AND KYRIAZIDOU, E. (2000): "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica*, 68, 839-874.
- HONORÉ, B., AND TAMER, E. (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 74, 611-629.
- HU, Y., AND S. M. SCHENNACH (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76 (1), 195-216.
- HU, Y., AND M. SHUM (2012): "Nonparametric Identification of Dynamic Models With Unobserved State Variables," *Journal of Econometrics*, 171, 32-44.
- HURWICZ L. (1950): "Generalization of the concept of identification," In *Statistical Inference in Dynamic Economic Models*, ed. T.C. Koopmans. New York: Wiley.

- ICHIMURA, H., AND T. THOMPSON (1998): “Maximum Likelihood Estimation of a Binary Choice Model with Random Coefficients of Unknown Distribution,” *Journal of Econometrics*, 86, 269-295.
- ISHWARAN, H. (1999): “Information in semiparametric mixtures of exponential families,” *Annals of Statistics* 27, 159-177.
- KHAN, S. AND D. NEKIPELOV (2016): “Information Structure and Statistical Information in Discrete Response Models,” unpublished manuscript.
- KHAN, S. AND E. TAMER (2010): “Irregular Identification, Support Conditions, and Inverse Weight Estimation,” *Econometrica*, 6, 2021-2042.
- KOOPMANS, T.C. (1949): “Identification Problems in Economic Model Construction,” *Econometrica*, 17, 125-144.
- KOOPMANS, T. C., AND O. REIRSOL (1950): “The Identification of Structural Characteristics,” *Annals of Mathematical Statistics*, 21, 165-181.
- KRESS, R. (1999): *Linear Integral Equations*. Springer.
- JOHNSON, E. G. (2004): “Identification in Discrete Choice Models With Fixed Effects,” Working Paper, Bureau of Labor Statistics.
- LE CAM, L. AND YANG, G. L. (1988): “On the preservation of local asymptotic normality under information loss”, *The Annals of Statistics*, 16, 483–520.
- LEWBEL, A. (2016): “The Identification Zoo – Meanings of Identification in Econometrics” Forthcoming in *the Journal of Economic Literature*.
- LUENBERGER, D. G. (1997): *Optimization by Vector Space Methods*. New York: John Wiley & Sons. 1969 edition.
- McFADDEN, D. AND K. TRAIN (2000), “Mixed MNL models for discrete response”, *Journal of Applied Econometrics*, 15, 447-470.
- MAGNAC, T. AND E. MAURIN (2008): “Partial Identification in Monotone Binary Models: Discrete Regressors and Interval Data,” *Review of Economic Studies*, 75, 835–864.
- MANSKI, C. (1975): “Maximum score estimation of the stochastic utility model of choice,” *Journal of Econometrics* 3, 205-228.
- MANSKI, C. (1987): “Semiparametric Analysis of Random Effects Linear Models From Binary Panel Data,” *Econometrica*, 55, 357-362.
- MASTEN, M.A. (2015): “Random Coefficients on Endogenous Variables in Simultaneous Equations Models”, working paper.

- MATZKIN, R. L. (2007). Nonparametric identification. In J.J. Heckman and E.E. Leamer (eds.), *Handbook of Econometrics*, Vol. 6b, pp.5307-5368, Elsevier, New York.
- MATZKIN, R. L. (2013): “Nonparametric Identification in Structural Economic Models,” *Annual Review of Economics*, 5.
- NEWKEY, W. K. (1990): “Semiparametric Efficiency Bounds,” *Journal of Applied Econometrics*, 5, 99-135.
- NEWKEY, W. K., AND J. POWELL (2003): “Instrumental Variables Estimation for Nonparametric Models,” *Econometrica*, 71, 1565–1578.
- RIDDER, G. AND T. WOUTERSEN (2003): “The singularity of the efficiency bound of the mixed proportional hazard model,” *Econometrica*, 71, 1579-1589.
- ROTHENBERG, R. J. (1971): “Identification in parametric models,” *Econometrica*, 39(3), 577-591.
- RUBIN, B. (1999): “Inversion and Characterization of the Hemispherical Transform,” *Journal d’Analyse Mathématique*, 77, 105-128.
- RUBIN, D., (1974): “Estimating causal effects of treatments in randomized and non-randomized studies,” *Journal of Educational Psychology*, 66, 688-701.
- RUDIN, W. (1973). *Functional analysis*. McGraw-Hill, New York.
- SANTOS, A. (2011), ‘Instrumental variables methods for recovering continuous linear functionals’, *Journal of Econometrics* 161, 129-146.
- SEVERINI, T. A., AND G. TRIPATHI (2006): “Some Identification Issues in Nonparametric Linear Models with Endogenous Regressors,” *Econometric Theory*, 22(2), 258–278.
- SEVERINI, T. A., AND G. TRIPATHI (2012): “Efficiency Bounds for Estimating Linear Functionals of Nonparametric Regression Models with Endogenous Regressors,” *Journal of Econometrics*, 170(2), 491-498.
- STINCHCOMBE, M. B. AND WHITE, H. (1998): “Consistent specification testing with nuisance parameters present only under the alternative,” *Econometric Theory* 14, 295-325.
- STEIN, C. (1956): “Efficient Nonparametric Testing and Estimation,” in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1. Berkeley: University of California Press.
- VAN DER VAART, A. W. (1991): “On differentiable functionals,” *The Annals of Statistics*, 19(1), 178–204.
- VAN DER VAART, A. W. (1998). *Asymptotic statistics*, vol. 3 of Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.